

Science on the electoral agenda?

The general opinion of the state of British science, provoked by last week's discussion of *Nature's* manifesto, is that morale matters more than money, but that much must be done to remedy the impoverishment of scientists.

LAST week began the season in which British political parties decamp for a week to seaside resorts (vacated by holiday-makers) to adopt resolutions on how they will behave in future: the Liberal Democrats met at Eastbourne. Predictably, nothing substantial was said about the party's policy on research and development. There were the usual references to the need for a skilled workforce, for improved education in primary and secondary schools and for easier access to higher education — and appropriate promises, "if in power", to bring about that state of grace.

That would be a case for not voting for the Liberal Democrats were it not that the conferences of the Labour (opposition) and Conservative (government) parties (yet to come) are unlikely to be much better. The Conservatives will congratulate themselves on deciding to merge the universities and polytechnics, establish a system of vocational qualifications for school-leavers and so on, while the Labour Party would have been advocating those steps as the linchpins of its science policy had not the government stolen its clothes. In the event, the chief difference between the two largest parties will be that Labour would appoint a science minister (in the prime minister's administrative entourage) and that the present government will not.

Last week's public meeting about this journal's *Manifesto for British science* showed that all this is a far cry from the anxiety of working scientists in Britain (see page 203). Naturally, nobody quarrels with the emphasis given to the education of the young in the plans of the several parties, but that is only half the battle. Three issues stand out from last week's meeting — pay (or, rather, personal impoverishment), the maintenance of the dual-support system and the place in a future scheme of things of Britain's three 'mission-orientated' research councils. In the weeks or months before the general election (which could be as soon as November), this journal will use its influence to see that note is taken of these issues.

The question of impoverishment is most easily dramatized by a numerical quantity — £4,100. This is the current annual stipend on which young people in Britain 'lucky' enough to win a studentship from the Science and Engineering Research Council (SERC) must keep body and soul together for the three or four years required to earn a PhD degree. The pressure group Save British Science was the first, earlier this year, to argue that this is a ludicrously small sum.

By serving as a public mark of how Britain values those whom all the political parties look to for a refashioning of the future, it is at least a disincentive to the choice of a career in research, and a demeaning one. But the same sum of money also explains how the nationally agreed starting salary for a fully qualified scientist in academic life should differ very little from that of a newly qualified secretary. The cost of putting this to rights would be so substantial that neither the government nor the opposition parties (all of whom seek reputations of fiscal probity) will commit themselves to reform. They should be nailed on this question.

The related issue of a career structure for researchers also requires attention. The notion that everybody should have lifelong tenure in his or her chosen field would not wash, and would also be counterproductive in the literal sense; the research enterprise needs constant renewal, which is predicated on mobility. The worry in Britain now is that, with the *de facto* abolition of tenure at universities, increasing numbers of those working as researchers in the academic sector do so on short-term contracts, and cannot enjoy even a reasonable sense of security about their futures. There are market solutions of this problem. Which party will be the first to seek them out?

Current anxiety about the dual-support system has its roots in the government's decision that British polytechnics should in future stand equally with the universities, financed by three common 'funding councils', and in the separate decision that, from the beginning of the next academic year, £150 million of the funding councils' money should be transferred to the research councils, enabling them to accompany research grants with overhead payments. Sir Mark Richmond, chairman of the Science and Engineering Research Council (SERC), courageously argued last week that this process cannot be allowed to go too far without robbing polytechnics and universities not now strong in research of the opportunity for self-improvement.

He is right. But is the government (or any of its rivals) willing to face the financial consequences? The worry is that tidy-minded civil servants will aim at converting the whole of the current research support in the higher education budget to overhead payments, and that political pressures (from Wales or Scotland, or on behalf of polytechnics generally) will then compromise the research councils' current criteria for the award of research grants.



Abolishing the divide between the two halves of the 'binary' system of polytechnics and universities has so far been an administrative business; neither the government nor the Labour Party (which would have done the same) has faced the truth that extra research support will be needed. (*Nature's* remedy for this state of affairs, a programme of mergers between individual institutions, would not obviate that need.)

The placing of the three mission-orientated research councils (for agriculture, medicine and the natural environment) in the future scheme of things is a different kind of issue. Many people last week were alarmed at the idea that these excellent organizations, after losing their present roles in the support of graduate education and academic research, should be converted into autonomous agencies and made directly dependent on the corresponding ministries. One fear is that they would then be neglected and, eventually, emasculated.

That fear springs from despair — the conviction that Britain will never be well governed. Nobody should underestimate the eagerness of civil servants to create a surprise-free environment for the ministers they serve, but is it entirely unreasonable to hope that, one day, government departments administering essentially technical spheres of public life (such as agriculture, health and the natural environment) will welcome a closer connection with basic research in their fields?

One of the shabbiest features of British research administration in the past few years is that the agriculture and health departments have walked away from their Rothschild obligations to the research councils — the former by repudiating them, the latter by handing over the funds specified to the Medical Research Council. But those who believe that basic research can illuminate the administration of technical spheres of public life (and who does not?) cannot but believe that agriculture, public health and the natural environment would be better administered if the departments concerned were required to listen to the outcome of the basic research they were compelled to pay for. Naturally, there would have to be a minister responsible (in part) for science to ensure that.

The abiding weakness of British research administration is the failure of ministers and their agents to appreciate what the research enterprise is about. Their ignorance is the more damaging because they are not aware of it. They would learn something from more effective coordination of the government's own huge spending on research, as well as from having to shoulder direct responsibility for spending on basic research. But it will be a long time before British governments will have changed so as to accommodate the most eloquent complaint at last week's meeting, that the incomprehension evident in British governments' demands of the research community has been in many ways more damaging than stop-go policies on various aspects of the budget.

How can these issues (and others) be made parts of the political agenda in the short time that may be left? There are old-fashioned ways of playing this trick. Researchers

might write to their Members of Parliament, for example. Britain's scientific societies should also chip in with public protests at the general maladministration of affairs with which they are vitally concerned. In short, there needs to be a fuss, even an undignified fuss. And soon. □

Indirect costs again

Data from a first round of audits hint that accounting problems may afflict nearly all US universities.

THE US research community continues under fire from Congress and government auditors, who have discovered that the complex accounting system by which universities charge overhead on federal grants has tolerated innumerable charges for items that have nothing to do with research. The most dramatic case so far is that in which Michigan congressman John Dingell revealed irregularities in costs for which Stanford University had been reimbursed — including part of the cost of a university yacht and fancy furnishings for the Stanford president's mansion. That scandal led in June to the resignation of university president Donald Kennedy (with effect from the end of this academic year) on the grounds that, being part of the problem, he could not be part of the solution.

So far, the scandal has centred on the most famous US private universities; in some quarters, people seem to have taken a particular pleasure in seeing the mighty brought low. But now, government auditors have discovered that the problem extends even to the University of Michigan in Dingell's home state. There, the auditors say, the federal government was charged for part of the cost of sending university officials from Michigan to sunny California — for the Rose Bowl football championship in 1989.

The apparent ubiquity of the abuses suggests that the problem is more with the system than with the individuals who operate it. There is no evidence, for instance, that Michigan University officials deliberately sat down with their business offices and said, "Let's steal money from research to pay for the Rose Bowl." Nevertheless, literally millions of dollars have been charged to research accounts because nobody took the trouble to segregate from administrative overhead costs items that are totally unrelated to the research enterprise.

Segregating costs is hardly beyond the capacity of good accountants. Dingell is accused by some scientists of running a vendetta against research in his persistent pursuit of allegations of research fraud. Whatever his motives there, on the indirect cost issue, he has uncovered a genuine scandal that deserves the serious attention of every research university in the United States. This is not a dispute about minor errors in a scientific manuscript, but one in which millions of dollars have been lost to scientific research. The disputed charges in Michigan alone add up to \$8.3 million. Dingell would do the research community a great service if he saw this issue through, especially if he then makes sure that the money recovered from improper charges is redirected to valid projects. □

AIDS scandal indicts French government

■ Blood test delayed to help French industry

■ Decision was made at the highest levels

CHOOSING economics and nationalism over public health, French government officials delayed the approval of a US AIDS test for five months in 1985 so that a French test could be released first, according to a report of a government investigation released last week.

The report, by Michel Lucas, the government inspector general for social affairs, was prompted by three years of lawsuits by the families of French haemophiliacs who had died of AIDS, and by the protest resignation last summer of Michel Garretta, director of the national blood transfusion centre. Based on internal documents and transcripts of government meetings, the Lucas report concludes that even though officials knew that haemophiliacs and other blood recipients would probably be infected by AIDS-contaminated blood supplies during the delay, they decided it was more important to prevent a US drug company from monopolizing the French test market than to speed the screening process by half a year.

Acting on instructions from Edmund Hervé, the secretary of state for health, the National Health Laboratory held up the blood test application of US-based Abbott Laboratories from 2 March 1985, when the test was approved in the United States, to 24 July, a month after a test by the French company Diagnostic Pasteur was approved, according to the Lucas report. Of the 3,000 haemophiliacs listed in France in 1985, 1,200 are now seropositive and 185 are dead. It will probably never be known how many of those were infected

during that five-month period.

Describing a trail of mistakes and deliberate subterfuge that Lucas suggests may lead all the way to the office of then-Prime Minister Laurent Fabius, the report reveals that government officials disregarded reports of contaminated blood supplies and growing infection statistics in order to give the test produced by Diagnostic Pasteur a chance to monopolize the domestic market.

According to a transcript of a meeting on 9 May of the cabinet, a representative of the health ministry noted that "the American firm Abbott and Diagnostic Pasteur [the marketing arm of the Pasteur Institute] are both ready to file an application to the national health laboratory. But it seems that Abbott is already looking at the national market and is starting an intensive promotion of their products at French blood transfusion centres."

"It is necessary to add that at this time that the American test has a considerable trump card; its price is half as expensive as the French product," the representative continued. "If the Diagnostic Pasteur test is not adopted in the blood transfusion centres now, it is clear there will not be a French market for the product and consequently there will not be an international market. A decision must be made rapidly, since the National Health Laboratory can not hold back Abbott's application very long beyond 13 May, the legal limit, without running the risk of contentious appeals."

Much of the emphasis on the econom-

ics — rather than the health risks — of AIDS can be explained by the fact that, even as late as mid-1985, French health officials were underestimating the severity of the threat. At the same 9 May meeting, the representative of the ministry of health noted that "cases of post-transfusion AIDS are generally rare — fewer than the number of post-transfusion hepatitis cases." Therefore, "the introduction [of a blood test] will not significantly affect the [spread of] the illness since only a few cases will be avoided. One is only going to arouse the worries of a lot of people ... without doing much to [halt] the evolution of the disease."

Officials were also concerned about the cost of screening all blood donations. At the meeting, the representative from the health ministry declared that he "was opposed to [covering] the price of the test with health insurance, because of the large sums involved." He estimated that on the basis of four million donors, at FF50 a test, the cost of the testing would be more than FF200 million (\$35 million).

In the report, Lucas suggests that officials were guilty of an intentional disregard for public safety, if not outright conspiracy. "I cannot believe that the [health officials] willingly distributed death, but it is necessary to recognize that they distributed products that they knew caused death in 10 per cent of all cases." In 1985, French scientists were still estimating that only 10 per cent of all seropositive individuals would develop AIDS.

Although some scientists tried to persuade the health ministry of the dangers of unscreened blood, "the importance of this message of alarm does not seem to have been perceived," Lucas notes.

The Lucas report will now go to current health minister Bruno Durieux, who is expected to turn it over to the justice ministry to decide whether to start legal proceedings. **Christopher Anderson**

"AIDSgate" — a chronology

9 January 1985 — In an internal memorandum to the health ministry, François Pinon, of the Cochin Hospital in Paris, warns that "the possibility of AIDS transmission by blood transfusion is certain."

11 February — Abbott Laboratories files an application for approval of its AIDS blood test in France, following a similar application in the United States.

28 February — Diagnostic Pasteur files for approval of its blood test in France.

2 March — Abbott's blood test is approved in the United States.

12 March — Jacques Roux of the Necker Hospital in Paris informs the health ministry that a review of the hospital's blood bank records has revealed that "a person with AIDS has given blood 20 times between 1979 and 1984." A few days later, the Paris-based

Cochin Hospital tells the health ministry that a search of its records has shown that 6 per cent of blood donations appear to have been made by people who have since been found to be carrying the AIDS virus. In an internal memorandum to the French director general of health, Cochin officials report that "it is probable that all the blood products prepared from pools of Parisian blood donors are currently contaminated."

25 April — Robert Netter, director of the French National Health Laboratory, writes to the secretary of state for health, "It doesn't seem possible to delay the American test much longer without risking an appeal [by Abbott] to the Council of State charging abuse of power. Under these conditions, I suggest granting an immediate approval for Pasteur and delaying Abbott to 13 May 1985" — the latest date that the

Laboratory could legally hold the application without taking action.

9 May — In a cabinet meeting chaired by François Gros, health adviser to then-Prime Minister Laurent Fabius, officials decide that the Abbott application should be delayed longer still. Michel Garretta, director of the national centre for blood transfusion, raises a solitary protest: "Given that three months of delay, more or less, means the death of 5 or 10 haemophiliacs...it is absolutely urgent to interrupt the illness now" by beginning a screening process. But the cabinet overrules him.

21 June — The Diagnostic Pasteur blood test is approved.

24 July — France finally approves the Abbott test, some five months after it was approved in the United States and six months after it was submitted to French authorities. **C.A.**

US science's 'stealth agency'

Washington

A SHADY White House office working under Vice President Dan Quayle is emerging as one of the biggest behind-the-scenes players in Washington science politics. Not surprisingly, that is starting to make a lot of people nervous.

The President's Council on Competitiveness is the successor to the Reagan-era Task Force on Regulatory Relief — and to the Task Force's controversial deregulation policies. Almost every government policy and decision is eligible to be sent through the Council for approval, if the White House sees fit.

The Council looks for policies in fields such as the environment, drug approval, risk assessment and laboratory regulation that may damage the country's economic health. Potentially, this could include nearly everything. In practice, the Council has chosen to make research fields one of its highest priorities, arguing that over-regulation and an excessive focus on government support in science and technology is particularly damaging to competitiveness.

For instance, at a meeting last week of the President's Council of Advisors on Science and Technology (PCAST), Allan Hubbard, executive director of the Council, said his staff is exploring ways to improve technology transfer at the national laboratories of the Department of Energy and the National Institutes of Health. "We're thinking that perhaps we should take away a little R&D money and require the labs to find private co-investors" to make up the difference, he said.

PCAST member John Foster, a former director of the Lawrence Livermore National Laboratory, thought it was a great idea. "We're spending too much on the labs," he agreed. "Let's cut their spending by 10 to 15 per cent to get their attention. Then take half of that and offer it back to them as long as they can find matching funds from industry." Hubbard promised to offer Foster's proposal to a Council working group on the issue.

Whether or not such a cut actually shows up in the national laboratory budgets, the episode illustrates the power the Council can wield.

Last year, after long negotiations with industry and activist groups, Congress passed a bill to make it more difficult for a drug to qualify as an 'orphan', one whose target audience is too small to justify research and development cost without special market protections. But on 9 November, President Bush vetoed the bill, catching almost everyone off guard. Who was behind the reversal? A fact sheet explaining the veto told the whole story with its letterhead alone — it came from the Council on Competitiveness.

Critics charge that the Council is essen-

tially a law unto itself. Operating in secrecy, free from Freedom of Information Act requests and answerable only to the President, the Council is a 'stealth agency', they claim. This month, for instance, the non-profit investigative group OMB Watch offered a dozen cases of Council 'interference' in environmental and health regulations, from reduction of wetlands standards to delays in quality-control reforms at clinical laboratories. "The Council has turned the term 'competitiveness' into a buzzword for eliminating regulations that corporations oppose," says OMB Watch executive director Gary Bass.

Industry takes a different view. Biotechnology companies showered the Council with praise in February, when it released a report calling for a reduction in the 'burden of regulation' on the industry.

Hubbard defends the Council as the last bastion of good sense in a bureaucracy that would regulate the country to an economic standstill, given the chance. "The role of the Council is to see that government is not interfering any more than is necessary in business," he told PCAST. "I don't see why cost-benefit analysis shouldn't enter into every regulation."

In the upcoming months, the Council will have even more opportunities to flex its muscle on:

■ **Food and Drug Administration (FDA)** drug approval. A council task force chaired by Former Health and Human Services deputy director Constance Horner is about to recommend major FDA reforms. One option that has been considered is privatizing the drug-approval arm of the agency by running it as a business based on user fees, much like the UK Medicine Control Agency. Other ways to streamline drug approval include 'fast tracking' more cancer and AIDS drugs, harmonizing US standards with those in other countries, including accepting data from foreign trials, and increasing the use of outside advisory bodies like the institutional review boards at many hospitals and universities.

■ **Biotechnology.** A Council working group is preparing to release a final version of its 'SCOPE' plan for biotechnology regulation. The new plan is expected to change existing regulation by setting levels of regulatory oversight based on the real risk inherent in the product rather than the process used to make it, an approach known as 'performance based' regulating.

■ **Clean Air Act.** The Council has been given the lead in turning last year's Clean Air Act into actual regulations. Hubbard says that the principal aim will be to reduce the measure's estimated \$25,000 million cost to US industry.

Christopher Anderson

US RESEARCH POLICY

A new role for DARPA?

Washington

SINCE the end of the Second World War, the US Department of Defense has been by far the leading supporter of research and development (R&D) in the United States, and even with the collapse of the Communist Bloc, that is unlikely to change soon. Nonetheless, as economic competitiveness becomes a more important part of national security and military superiority becomes less urgent, the US government may start to look to defence R&D for more than high-technology weaponry.

That, at least, is the suggestion of a report released last week by the Carnegie Commission on Science, Technology and Government, a private group comprising distinguished scientists, community leaders and former government officials. The report, which was the subject of a hearing by the congressional Joint Economic Committee, contains a variety of prescriptions for helping US economic performance, but the most controversial centres on the interplay between military and commercial R&D.

"The United States now has two technology bases, a defence technology base and a commercial technology base," said retired Admiral Bobby Inman, chairman

of the group that prepared the report. "The nation must move toward a single national technology base." To that end, the report says, the Defense Advanced Research Projects Agency (DARPA), which funds research into advanced technology with potential military applications, should be transformed into the National Advanced Research Projects Agency (NARPA). NARPA not only would fund military research but also would support research into advanced civilian technologies that federal agencies other than the defence department request.

DARPA, Inman noted, has a good record in funding 'dual use' technologies — those with both a military purpose and potential commercial applications. Reconstituted as NARPA, it could be expected to play a big role in promoting research with long-term commercial rewards.

The suggestion for a NARPA-to-DARPA transformation, however, has much more to do with assuring that the defence department has access to the best technology than in improving economic competitiveness, said Lewis Branscomb of Harvard University, a member of the group that prepared the report. When the military employs dual-use technology, it

often ends up with less advanced products than are available commercially, for a variety of reasons, including the rigidity of military specifications. "It is through partnerships in dual-use technology that defence has the best chance to gain access to leading commercial technologies," Branscomb said.

Inman acknowledged that the defence department will not be happy with allowing other government departments to have a say in the direction of DARPA, which would be inevitable if it were transformed into NARPA. But, he added, "the bulk of the opposition at the Department of Defense to NARPA arises from the sense that it will take money away from defence research and give it to commercial research." That is not the intention of the report's recommendation, he said.

Murray Weidenbaum, an economist from Washington University in St Louis, argued that expanding NARPA's mission to include civilian technologies would damage its effectiveness by asking it to do too many things. To make sure that the military gets the best dual-use technology, it is not necessary to change DARPA; instead, "the Pentagon should reduce the obstacles to its procurement of state-of-the-art products available in commercial markets."

The Carnegie group would also include economic and technology issues in the purview of the National Security Council (NSC), the group that advises the president on national security matters. As the NSC has regular access to the president, this should help make economic and technology policy a priority for the president.

The recommendations of the Carnegie Commission are unlikely to be embodied as legislation; at present, they are little more than food for thought. But they do clearly illustrate an issue that is likely to grow in importance in the United States over the coming years. As the threat of war with the Soviet Union and its erstwhile allies recedes, the major threats to US national security are likely to be economic ones. This in turn will greatly change the way in which science and technology research is funded, and the struggle to determine how those dollars are spent will not be over soon.

Robert Pool

CORRECTION

The first sentence in the news article "Cold fusion tempest at MIT" (*Nature* 353, 98; 12 September 1991) should have read: "A former Massachusetts Institute of Technology (MIT) official has requested a scientific misconduct inquiry at the university's Plasma Fusion Center, claiming that researchers there manipulated data and the MIT press office in conducting an 'orchestrated attack' against cold fusion and its advocates." Because of an editing error, the original sentence incorrectly represented the MIT official as claiming that the press office, rather than the researchers, attacked cold fusion.

Primate import probe

A MIAMI grand jury and the US Fish and Wildlife Service are investigating charges that the president of one of the largest US primates import companies brought Asian orangutans caught in the wild into the United States, contrary to US law. Earlier this month, the US District Court in Miami issued several subpoenas for documents pertaining to the import activities of Matthew Block, president of Miami-based World Wide Primates, which imports about a quarter of the 12,000 to 14,000 research primates brought into the country each year. This is not the first time Block has attracted controversy, particularly from the International Primate Protection League (IPPL) and its chairwoman, Shirley McGreal. In a statement released early this week Block said that any investigation "which derives from McGreal's false allegations is an outrage. It may well be that United States authorities have been misled by McGreal and her organization." Last August he filed suit against McGreal for damaging his professional reputation after she distributed copies of a 1990 Centres for Disease Control inspection report on his facilities together with a brief letter asking recipients to consider if they should be doing business with him.

C.A.

Satellite lent to US

DOWN to one ageing weather satellite with five more nowhere near ready to go, the United States last week turned to Europe for a loan. An agreement between the US National Oceanic and Atmospheric Administration (NOAA), the European Space Agency, and the European Meteorological Satellite Corporation will move the European Meteorsat-3 geosynchronous satellite 48 degrees west from its current position over South America so that it can cover the lower 48 US states. The lease of the European satellite should ensure that NOAA has at least one weather satellite operating at all times until the end of 1993, when the first of the next generation of weather satellites — known as GOES-NEXT — is expected to be launched. Budget overruns and technical difficulties have conspired to delay the GOES-NEXT programme by at least three years (see *Nature* 352, 362; 1 August 1991).

C.A.

Sex study dead again

A PROJECT to survey Americans on their sex habits so that researchers can better understand the spread of sexually transmitted diseases has been rejected again, this time by conservative senator Jesse Helms (Republican, North Carolina). During last week's full Senate vote on the 1992 Health and Human Services (HHS) Appropriations bill, in which \$10 million was included for both an adult and a teenage sex survey by the National Institutes of Health (NIH), Helms won strong support for an amendment to transfer the funds out of NIH and into a pro-abstinence project known as the Adolescent Family Life

Program. This is more unfortunate for the NIH than it may sound. HHS secretary Louis Sullivan had already cancelled the \$8 million adult survey earlier this year after it was initially approved. That money was to return to the NIH general research fund; now it will leave NIH entirely.

C.A.

Ozone mission up

INAUGURATING its next generation of science missions — the Earth-observing 'Mission to Planet Earth' probes — the US National Aeronautics and Space Administration last week launched the Upper Atmosphere Research Satellite (UARS) with only one minor hitch. Astronauts aboard the space shuttle Discovery deployed the \$740-million satellite last weekend to begin its 26- to 28-month planned mission in a circular near-polar orbit at an altitude of 600 km. A transponder used to receive data from Earth stations failed to work properly, but ground controllers were able to switch to a backup receiver. Four of the ten instruments aboard UARS will measure gases — including CFCs, nitric oxide and chlorine monoxide — thought to be responsible for ozone depletion in the upper atmosphere, while other instruments will measure the intensity of the Sun's ultraviolet radiation, solar particles, and atmospheric temperature and winds.

C.A.

Software protection

SEEKING to spur computer research collaboration between US government and industrial researchers, Senator John Rockefeller (Democrat, West Virginia) wants to keep government software from automatically falling into the hands of the public. He has introduced a bill that would make an exception to current copyright law (which now makes government-developed software freely available to all) for the products of cooperative research and development agreements between federal laboratory and industrial partners. Rockefeller hopes that this will reassure companies that their intellectual property will be protected. But his bill has run into opposition from the American Civil Liberties Union (ACLU). Joined by such mainstream computer groups as the Information Industry Association and the Institute of Electrical and Electronics Engineers, the ACLU testified last week that copyrighting federal software would violate the public right of access to government information.

C.A.

AIDS meet relocates

THE eighth international AIDS meeting will be held in Amsterdam next July, sponsors of the meeting announced last week. The meeting had originally been scheduled for Boston, but its organizers decided to move it because of US restrictions on visitors infected with the human immunodeficiency virus (see *Nature* 353, 96; 12 September 1991). Special-interest groups representing people with AIDS had threatened to disrupt the meeting if it were held in the United States while the travel restrictions were still in place.

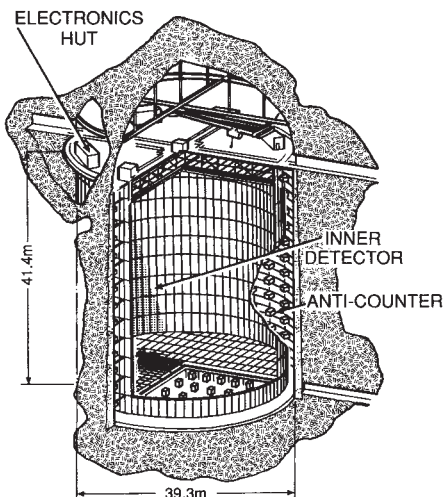
R.P.

Bigger than 20 Olympic pools

Tokyo

AFTER years of lobbying by Tokyo University researchers, the Ministry of Education, Science and Culture finally submitted a budget request at the end of last month for a giant neutrino detector to be built in a mine in Gifu Prefecture. Super-Kamiokande, which is expected to be completed in 1996, will be the largest and most sensitive observatory for neutrinos and nucleon decay in the world.

The device is a scaled-up version of the Kamiokande detector which made headlines in science in 1987 by detecting neutrinos from the supernova in the Large Magellanic Cloud. The new detector, which will be more than ten times the size of Kamiokande, will contain 50,000 tons of water — enough to fill more than 20 Olympic swimming pools — and it will be able to detect and trace the source of neu-



The planned Super-Kamiokande detector trinos from supernovae and even to pinpoint supernovae too faint to be detected by optical telescopes. Furthermore, with fast on-line analysis by computer, Super-Kamiokande will be able to locate supernovae and inform optical observatories of their position before the exploding stars become bright, say researchers at the Institute of Cosmic Ray Research of Tokyo University who will build the detector.

Super-Kamiokande should also be able to pick up the relic cosmic particles from formation of the early Universe, such as relic neutrinos from past supernova. And it may even detect the formation of black holes from the abrupt disappearance of neutrino bursts when a neutrino signal is cut-off by black hole formation, the Tokyo researchers say.

The detector works by picking up Cerenkov radiation from high-energy electrons produced, for example, when a neutrino enters the tank or when a proton in the water decays. The inner walls of the tank will be lined with 11,200 highly sensitive photomultiplier tubes to detect and

locate the radiation. Another 700 photomultiplier tubes in the outer wall will help to filter out background radiation (see figure).

The tank will be encased in solid rock 1 km underground to screen out unwanted cosmic radiation. And it will be filled with water from the mine, which, after filtration, will be sufficiently transparent for the photomultiplier tubes to pick up tiny bursts of Cerenkov radiation even from the centre of the gigantic tank.

The education ministry has applied for ¥613 million (\$4.5 million) in next fiscal

JAPAN

Is a giant linac next?

Tokyo

Now that one group of physicists at Tokyo University has won funding for the giant neutrino observatory Super-Kamiokande (see above), a second group is expected to push even harder for its plan to build a huge new accelerator complex equipped with a 1-GeV proton linear accelerator (linac). If approved, the Japan Hadron Project will not only cost hundreds of millions of dollars, it will also involve a major reorganization of Tokyo University and could lead to the formation of a new national research institute to run the facility.

The project has been promoted for several years by the university's Institute for Nuclear Study. Preliminary studies and committee meetings to discuss the project were supported this fiscal year with ¥4 million (\$30,000) from the Ministry of Education, Science and Culture. The ministry is expected to provide more funds next year for an in-depth study of the project and the related reorganization of Tokyo University that it would require.

Under the institute's plan, a 1-GeV proton linear accelerator would be built next to the High Energy Physics Laboratory (KEK) in Tsukuba, which houses the huge positron-electron collider TRISTAN and other accelerator facilities. As well as feeding 1-GeV protons to an existing 12-GeV proton synchrotron in KEK — an arrangement that would substantially improve the quality and intensity of the KEK synchrotron beam — the proton accelerator would supply beams of protons to a meson science facility and to a neutron scattering facility. The protons would also be injected into a heavy-ion linear accelerator that would produce beams of exotic nuclei.

The great variety of beams available in the proposed complex would make possible a wide range of experiments in condensed-matter physics, chemistry and the life sciences, as well as in particle and

year's budget to cover the costs of the first 2,000 of the huge 20-inch photomultiplier tubes. Hamamatsu Photonics, a small Japanese company, is the world's only manufacturer of such tubes, which must be meticulously put together by hand.

The education ministry estimates that the overall cost of Super-Kamiokande will be about ¥8,700 million (\$64 million). Funds for excavation of the gigantic hole for the detector have already been approved in this year's budget. And Mitsui Mining and Smelting Company, which owns the mine in Gifu Prefecture, will begin preparatory work for the excavation next month.

David Swinbanks

nuclear physics. And proponents of the scheme are calling for the creation of a new 'national research institute for joint university use' to operate the accelerator complex.

Such national research institutes are among the best-funded organizations run by the education ministry, and they include KEK and the Institute of Space and Astronautical Science. The Institute for Nuclear Study has been lobbying hard to be designated as a national research institute. If it succeeds, it will follow in the footsteps of the National Astronomical Observatory and the Institute of Space and Astronautical Science, both of which were part of Tokyo University.

But there is some resistance to such a move within other parts of the university. Among the contentious questions to be examined next year are: Who will teach nuclear studies at Tokyo University if the institute separates from the university? What effect will the move have on the closely related Institute of Cosmic Ray Research, which branched off from the institute in 1976? Senior university officials want to ensure that the new facility maintains close links with the university, and it is still unclear how the proposed new institute would interact with KEK next door.

The total cost of the new complex is variously estimated at ¥30,000 million to ¥40,000 million (\$220-300 million), excluding the cost of the building to house the accelerators. Construction of related facilities attached to the 12-GeV synchrotron at KEK has already begun. A few years ago, Tokyo University decided to give priority to the Super-Kamiokande project in its budget negotiations with the education ministry (*Nature* **331**, 379; 1988), and now that Super-Kamiokande is under way, a push to launch the Japan Hadron Project can be expected. But approval is unlikely for at least a year or two.

David Swinbanks

Coping with a talent flood

Jerusalem

WITH immigrant engineers and PhDs pushing brooms on the streets of Tel Aviv suburbs, the term 'sanitary engineer' has lost its tongue-in-cheek character when referring to street cleaners in Israel. The country has not yet been able to cope with the flood of immigrant scientists and engineers sweeping in from the Soviet Union.

But in the face of this case of talent overload, the Israeli government and private foundations are trying a variety of strategies to integrate all this brainpower into the country's economy. If the effort is successful, the newcomers could trigger a new era of economic growth for the country.

Since the current wave of mass immigration began in December 1989, some 350,000 Soviet Jews have arrived in Israel, including 4,500 scientists and almost 40,000 engineers. In all, one million Soviet immigrants are expected to arrive in Israel in a five-year period, increasing the country's population by some 20 per cent.

Even if the reforms in the Soviet Union cause the tide to slow, Israel will still face a major adjustment. "The burden is overwhelming," says Deborah Lipson of the Zionist Forum, a volunteer organization for Soviet immigrants. "It is the equivalent of 15 million immigrants arriving in the United States in one year."

With the burden, however, comes an opportunity. Because Israel has few significant natural resources, its fortunes rest in large part on its scientific and technical skills, and the anticipated immigration should double the number of scientists and quadruple the number of engineers in Israel by 1995.

Aware of the potential, the government has in the past year started to nurture 2,000 immigrant scientists and engineers through their first years in the country. Some \$70 million has been allocated to programmes aimed at providing interim employment for these newcomers until they can be absorbed by the country's expanding industrial and academic base.

Many of the immigrant scientists and most of the engineers, however, must cope on their own. For most this means temporary employment in less exalted fields — some have actually become municipal street cleaners — and many face a permanent change in career.

"The problem is not so much the numbers of scientists and engineers, it's the condensed time span in which they're coming," says Baruch Eyal, a Ministry of Science official dealing with immigrant scientists. An earlier immigration wave that brought 180,000 newcomers from the Soviet Union beginning in the 1970s saw an even higher percentage of scientists, he

noted, but the flow had been spread over 15 years and appropriate jobs were found for virtually all of them. "The absorption of scientists and engineers depends on the economic growth of the country and unfortunately the growth in immigration now is much higher than the growth of the GNP," he said.

The most ambitious government programme is the creation of technological incubators around the country in which immigrant scientists and engineers work on projects with commercial potential that have been proposed by local industries or that the new citizens have proposed themselves. (By one estimate, Soviet immigrant scientists and engineers have brought with them some 100,000 patents.) Fourteen such centres are to be set up — several are already functioning — and they will provide employment for at least a year or two for some 700 immigrants, mostly PhDs. The hope is that the incubators will come up with new processes or products that will draw private capital.

In setting up the centres, the government is giving preference to fields that Israel is interested in developing, such as microelectronics, electro-optics, chemistry, software and superconductivity. The incubators are in existing universities, research institutes and industries.

Some immigrant scientists are finding jobs directly in industry. To encourage this process, the (Immigrant) Absorption Ministry pays about 85 per cent of the modest basic salary (about \$800 a month) for the first year in the hope that their employers will take them on as unsubsidized workers thereafter.

A fund administered by the Israel Academy of Sciences provides special treatment for a score of scientific superstars among the immigrants, providing them with comfortable salaries and laboratories for an extended period.

Israel's seven institutions of higher education have provided a limited number of employment opportunities for immigrant academics who have completed the obligatory six months of intensive Hebrew language courses. An expected expansion of universities — largely because of the Soviet immigration — will provide some 2,000 faculty posts over the next five years, and half are expected to go to immigrants.

Many of the immigrants have adjusted to the need for a change of career. Scores of Soviet science PhDs are being retrained as high school teachers, which should help improve science education in Israel. Immigrant geologists are a glut on the market and many are being retrained as surveyors. Some social scientists have registered for retraining as social workers. Marxist-trained economists will have to

make radical career adjustments, but Soviet psychologists surprisingly find it easy to find work in Israel after mastering the language. "The first Soviet psychology faculty was not established until 1966 in Moscow and it was set up along Western lines," says Shmuel Adler, who heads the department dealing with scientists at the Absorption Ministry. "But we have more trouble finding acceptance for Soviet-trained psychiatrists."

Israeli administrators have found the quality of the Soviet scientists and academics uneven. "In the theoretical sciences — math, physics, astrophysics — they're at the top of their field," Adler says. However, many of the other scientists and engineers must learn to adjust to Western standards and Western equipment.

The placement of immigrant scientists has also become the object of a number of private and institutional efforts. The Satec Corporation was founded for this purpose by an immigrant Soviet scientist, Herman Branover, who arrived in Israel in 1972 and now teaches magnetohydrodynamics at Ben-Gurion University in Beersheba. Branover established the high-technology company in Jerusalem four years ago with funding from foreign investors. It now employs 15 Soviet immigrant scientists and engineers together with sales and administrative staff.

One of Satec's scientists is Solomon Flax, a 60-year-old hydro-metallurgist who developed a technique in the Soviet Union for removing precious metals from industrial wastes using an environment-friendly and relatively cheap bromide solution rather than the cyanide conventionally used. Although he held more than 40 patents in the Soviet Union, he had not patented this process before he was dismissed from his university research job in the Ukraine after applying for an emigration visa to Israel. For nine years, barred from research work, he supported his family by using his technique to extract precious metals from castoff objects and then selling the metal. At Satec, he has obtained a US patent on his process and is negotiating with foreign companies for a pilot project.

Government planners hope that immigration will help to increase the gross national product by 10 per cent in 1992 and that the expansion of the economy by 1993 will permit absorption of as many immigrant scientists as arrive that year. By 1994, they hope to be able to deal with the backlog of unemployed scientists. Whether or not this is a realistic estimate, the transition period will clearly be painful for many of the immigrants, as well as a major burden for the country as a whole. If all goes well, however, the Israeli economy will be well prepared for the new millennium.

Abraham Rabinovich

Help from an unusual source

Tokyo

HIGH-TECHNOLOGY military equipment has found an unexpected role in helping scientists to study the erupting volcano Mount Unzen, which discharged again Sunday, destroying about 100 homes and a school.

After a pyroclastic flow of hot volcanic gas and ash swept down the slope of Unzen on 3 June, killing more than 40 people, a troop of about 600 self-defence force personnel were sent to the area, initially to search for survivors and bodies in the disaster zone. But soon after arrival, the army set up a variety of devices to detect pyroclastic flows from the volcano, and the self-defence forces now play a vital role in the scientific monitoring of Unzen.



Tanks for the help

The expensive equipment deployed by the troops far outshines the facilities at nearby government-run observatories. For example, the latest high-technology Doppler radar, designed for detecting attacking jet fighters and missiles, has been trained on the erupting volcano to pick up the high-speed pyroclastic flows, which at any time could come rushing down Mount Unzen at speeds of hundreds of kilometres an hour.

In addition, about ten soldiers are permanently deployed at the local Shimabara

Earthquake and Volcano Observatory of Kyushu University. They maintain a 24-hour watch for possible seismic activity associated with the volcanic eruption and are in constant touch with military headquarters to report any evidence of pending volcanic activity. The soldiers have become so expert in their interpretation of seismographs in the past three months that they sometimes substitute for the limited number of scientists at the observatory during the small hours of the night. Kazuya Ohta, director of the observatory, says he would welcome the soldiers as assistant researchers any time.

The military also frequently flies scientists over the volcano in self-defence helicopters to check the condition of the huge lava dome that sits precariously on the edge of the volcano's crater. And soldiers in Shimabara, who outnumber scientists by about ten to one, occasionally venture into the declared danger zone at the foot of the volcano in their armoured vehicles and tanks to collect samples for the researchers.

Unzen provides a rare opportunity for Japan's volcanologists to watch pyroclastic flows in action. The observations should yield valuable data for the interpretation of such flows in the geological record.

The lava dome of Unzen continues to grow. In the past few weeks, pyroclastic flows have headed in new directions, forcing the evacuation of another 1,000 or so people and bringing the total number of evacuees to more than 10,000. There is no sign that these people, many of whom have been without a home for more than three months, can return to their former homes in the near future. Even with the presence of the Japanese military, the danger of a 'surprise attack' by the volcano remains too great.

David Swinbanks

EARTH MAPPING

New Indian satellite is up

New Delhi

INDIA'S efforts to harness space technology for managing its natural resources took a leap forward on 29 August when its second remote-sensing satellite, IRS-1B, went into orbit after a successful launch from Baikonur Cosmodrome in central Asia. The launch took place as scheduled despite fears in India that it might be postponed because of internal problems in the Soviet Union following the abortive coup.

Designed and built by the Indian Space Research Organisation (ISRO) in Bangalore, the 980-kg polar orbiting satellite is identical to its predecessor IRS-1A, launched in March 1988. IRS-1A has

mapped all of India for ground water potential, and has also made a soil map, a land use map and a forest map. The launch of IRS-1B will double the country's capacity for data acquisition. Together, the two spacecraft will be able to cover the entire subcontinent once every 12 days as compared to 22 days by IRS-1A alone. ISRO also hopes to sell the images to other countries in the region.

ISRO says its future remote-sensing satellites IRS-1C and IRS-1D will have better resolution, stereo viewing, on-board recording and more frequent revisit capabilities. These are planned to be launched in 1993 and 1996.

K.S. Jayaraman

Chasing the source of radioactive steel

New Delhi

INDIA'S Atomic Energy Commission is investigating the circumstances that resulted in radioactive contamination of steel products exported to the United States by a private company in Calcutta. The inquiry follows last week's discovery by the US Nuclear Regulatory Commission (NRC) that the products contained traces of radioactive cobalt-60.

The steel products included several thousand tons of chain links for fencing, gate bars and truss rods. They were imported from Katia Steel Rolling Mills in Calcutta by Master Halco Inc., a US distributor in La Habra, California, and had been redistributed to retailers and wholesalers throughout the United States. According to the NRC, 5 per cent of the chain links tested were slightly radioactive. Further shipments from India have been halted, and the NRC has contacted the Indian government to ensure that it is aware of the problem and to determine the scope and source of contamination.

The Bhabha Atomic Research Centre (BARC) in Bombay is the only place in India that produces cobalt-60 in large amounts for industrial, medical and agricultural applications. But BARC has not supplied cobalt-60 sources to any steel rolling mills in the country and the chairman of the atomic energy commission, P.K. Iyengar, ruled out BARC-made cobalt-60 as being the source of contamination.

Iyengar says the cobalt-60 could have originated outside India because the Calcutta company was making steel by mixing locally available as well as imported scrap iron. This practice is common among steel rolling mills in India because imported scrap is cheaper. "Our suspicion is that the imported scrap used in steel-making might have been contaminated with cobalt-60," Iyengar said. Most blast furnaces abroad — not in India — use cobalt-60 in detectors to check the lining of the furnace for erosion. According to Iyengar, if the cobalt source accidentally falls inside the furnace, all the iron will be contaminated. The Atomic Energy Commission team will analyse the samples of scrap iron in the possession of the Calcutta company to find out whether or not the imported scrap is contaminated.

Some 25 years ago, aluminium pipes sold in 'thieves' bazaars' in Bombay were found to be radioactive. Investigations revealed that the pipes had been fashioned out of material removed from a radioactive waste dump in one of the BARC facilities. But BARC sources insist that present security arrangements are so tight that no waste material can escape from the grounds.

K.S. Jayaraman

Richmond issues challenge

London

"A NATIONAL disgrace" is how Sir Mark Richmond, chairman of the Science and Engineering Research Council (SERC), described the low pay of junior scientists in Britain, in an address to launch *Nature's Manifesto for British Science* last week (Friday, 13 September). Speaking in a personal capacity, Richmond challenged Kenneth Clarke, the Secretary of State for Education and Science, to increase scientific salaries sharply and immediately or face charges of "anti-intellectualism".

"Certainly the government seems to see no electoral disadvantage in denying our research workers the rewards their intellectual gifts are due," he said. "The Secretary of State ... gives every indication of priding himself on his qualities as a minister. ... Let him get the resources to provide the opportunity and the remuneration needed to make a decision by a person to go into a career in science no longer an act of total self-immolation."

But salaries are just one aspect of British science eroded by successive governments. More insidious is the steady attrition of funds experienced by the Universities Funding Council (UFC) and the research councils, the two parts of the 'dual support system' that underwrites most of British academic science.

In this system, the government (through the UFC) provides research facilities, while the research councils inject funds for particular projects. The system has been a success, but during the 1980s researchers have found it hard to make ends meet. Richmond blames the "failure" of the government to account for inflation in its

allocations to the UFC.

The plight of university research has been masked to some extent because research councils have taken up some of the slack. Although the number of teaching staff in universities remained more or less constant throughout the 1980s, the value of alpha-rated proposals for research council grants more than doubled, from £30 million in 1980 to a peak of more than £70 million in 1987, a demand that the research councils have been progressively less able to meet.

The research councils are also being asked to meet the costs of overheads — something that universities, through the UFC, should be doing. Having done for the universities, the government has focused its attack on the research councils themselves — a double blow to the dual support system. 'Stop-go' funding has led to much unnecessary disruption as the research councils have struggled to meet their various obligations.

Some have argued that the dual support system should be abandoned in favour of research-council-funded institutes 'at one remove' from universities. This would be "disastrous", says Richmond, because it would deny undergraduates the opportunity to move smoothly into research: "It is impaired opportunity which is one of the most baleful influences on scientific research in our universities at present."

Richmond argues for an improved dual support system in which the costs of research and teaching are properly accounted for as well as for the splitting of the recently merged UFC and PCFC (Polytechnics and Colleges Funding Council) into

Teaching (TFC) and Research (RFC) bodies (see figure). University research money and research council funds would be channelled through graduate schools that each university might be obliged to create as a separate 'cost centre' before it could receive funds for research.

In the meantime, the Advisory Board for the Research Councils (ABRC) would be replaced by an Advisory Board for Research (ABR) that would advise the Secretary of State on matters of policy, but otherwise let the research councils get on with the job.

But without the realization in government that the research base is a cultural asset, such schemes are rather like rearranging the deck-chairs on the *Titanic*. An extra £100 million annually to oil the wheels would lift flagging morale, quite apart from a boost in the total cost of research salaries by £130 million — just to begin with.

But Richmond takes issue with *Nature's* Manifesto in his support for the diversity of views fostered by several different research councils. *Nature's* proposal to scrap all except SERC and incorporate the others as agencies of government departments came under fire from another quarter; speaking from the floor, Sir Gordon Cox, the 85-year-old ex-secretary of the then Agricultural Research Council noted that the research council idea was proposed (by the Haldane Report of 1918) specifically as a defence against meddling civil servants.

These views were echoed by Jeremy Bray, the Labour Party's spokesman on science, who was the only representative from a political party present. He criticized the Manifesto's recommendation for a new government to spend in its first year "only what the present planning figures allow for" as "hardly a clarion call".

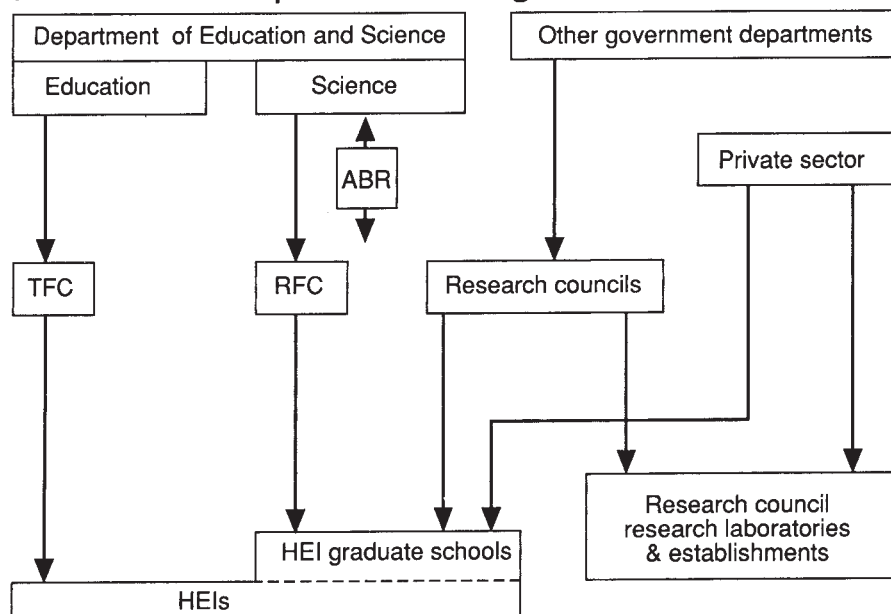
Science, he said, will never be high on the political agenda so long as cash calls are seen as "jobs for the boys". The science lobby should emphasize the value of its output, rather than its need for input. In the runup to a general election, scientists should ask their prospective political representatives a few difficult questions.

Personal stories from the 'shop floor' illustrating the economic depths to which young researchers must sink in pursuit of their goals were particularly telling.

One speaker said there is no guarantee that young researchers, fresh from their PhDs, will be able to secure even the pitifully low salaries on offer for any length of time. Perhaps more than money, researchers crave some kind of career structure. At present, the career of a scientist is as secure as that of an actor, except that the pay is even worse. Scientists, as Richmond noted, do science for the love of it, but should they "be abused by taking advantage of their obsession to pursue good science"?

Henry Gee

Sir Mark Richmond's plan for research organization in the UK



ABR=Advisory Board for Research; HEI=Higher Education Institution; RFC=Research Funding Council; TFC=Teaching Funding Council

Imaginary hazards?

SIR — A recent leading article ("Ethical problems to merit the name", *Nature* 352, 359–60; 1991) gave a provocative, profound and largely accurate perspective on the 'industry' consisting of committees performing ethical enquiries pertaining to the new biology. Correctly, it took them — and implicitly, society — to task for routinely defining contemporary ethical dilemmas as novel while neglecting essential corollaries to earlier controversies that have been addressed and reconciled. However, in one small respect, we would argue that the article has perpetrated just what it decries, by implying that any possible "environmental consequences if [genetically] altered organisms are more generally released" are somehow unique to or are more likely with the new biology. Given the ubiquity and impressive safety record of genetically improved — even 'transgenic' — organisms and the fundamental equivalence of old and new techniques of genetic manipulation, one is led to wonder, more generally released than what?

Two landmark documents from the US National Academy of Sciences (NAS) and the National Research Council (NRC) yield important scientific insights into this question. The NAS white paper¹ concluded that (1) "there is no

evidence of the existence of unique hazards either in the use of rDNA techniques or in the movement of genes between unrelated organisms"; (2) "[t]he risks associated with the introduction of rDNA-engineered organisms are the same in kind as those associated with the introduction of unmodified organisms and organisms modified by other methods"; and (3) "[a]ssessment of risks associated with introducing recombinant DNA organisms into the environment should be based on the nature of the organism and of the environment into which the organism is to be introduced," and independent of the method of engineering *per se*. In a more comprehensive treatment of this subject², the NRC committee agreed that "no conceptual distinction exists between genetic modification of plants and microorganisms by classical methods or by molecular techniques that modify and transfer genes," whether in the laboratory, in the field, or in large-scale environmental introductions. The report emphasized the vast experience gained from routine field trials of plants modified by these older genetic methods, noting that "an individual corn, soybean, wheat, or potato breeder may introduce into the field 50,000 genotypes per year on average or 2,000,000 in a career." And as reported

by Goodman *et al.*³, many crops commonly grown and consumed in the United States — including oats, tomatoes, potato, wheat, rice and maize — have resulted from interspecific or intergeneric gene transfer, yielding 'genetically altered' and 'transgenic' plants by any reasonable definition. Hence, we return to the question, more generally released than what? Than the hundreds of millions of acres already planted with 'transgenic' (but non-'recombinant') plants, perhaps.

We suggest that some concerns about research or commercial introductions of organisms — such as the use of pathogenic or invasive species — are justified but are not new or unique to the new biology and that, in fact, organisms manipulated with the newest, most precise, most predictable techniques are even *less* likely to cause problems. Application of the new biology is not ethically daring, to use *Nature's* phrase, but logic and common sense applied to society's scientific questions seem increasingly to be.

HENRY I. MILLER

Office of Biotechnology,
Food and Drug Administration,
Rockville, Maryland 20857, USA

SUSANNE L. HUTTNER

University of California,
Systemwide Biotechnology Program,
University of California, Los Angeles,
Los Angeles, California 90024, USA

1. Introduction of Recombinant DNA-Engineered Organisms into the Environment: Key Issues (National Academy Press, Washington, D. C. 1987).
2. Field Testing Genetically Modified Organisms: Framework for Decisions (National Academy Press, Washington, DC, 1989).
3. Goodman, R. M., Hauptli, H., Crossway, A. & Knauft, V. C. *Science* 236, 48–54 (1987).

OSI procedures

SIR — In her recent article (*Nature* 352, 563; 1991), Barbara Culliton writes that "Hadley . . . has explicitly told scientific groups that she finds intent irrelevant in identifying misconduct". I have not made this assertion, nor do I believe that it is true. What I have said, based on counsel of the National Institutes of Health (NIH) Legal Advisor, is that proof of intent is not *necessary* to a finding of scientific misconduct; the commission of acts proscribed by Public Health Service (PHS) regulations is itself sufficient for a finding that scientific misconduct has occurred.

She asserts that "Hadley's organizing principle for the OSI" (Office of Scientific Integrity) was to "keep . . . allegedly damaging data secret, available to the accused only in summary form". This assertion is flatly incorrect. I always followed PHS Policies and Procedures (*Federal Register* 56, 13 June 1991, page 27388), which state that subjects of an inquiry or investigation "are provided access to any research data under review . . . [and] are provided with an opportunity to review and comment on significant investigatory documents . . .".

Culliton also asserts that the "OSI's definition of due process . . ." expressly denies the accused the right to see at first hand all of the evidence against him or her". As I have said, this assertion is not correct. The only instance of which I am aware in which there was any departure from PHS policy occurred when the OSI did not have possession or control of certain pieces of evidence.

Culliton states that NIH Director Bernadine Healy "is distressed that Hadley . . . blocked the early release of the Gallo report to an advisory committee that was not duly constituted". Even Healy has not made such an allegation, and in any case it is entirely false.

SUZANNE W. HADLEY
(Special Assistant,
Office of the Director)

Office of Science Policy and Legislation,
National Institutes of Health,
Bethesda, Maryland 20892, USA

■ Barbara Culliton writes: Why, if damaging data are so faithfully provided, do so many scientists and their lawyers complain of secrecy?

But on the last point, Hadley is correct: the release of the Gallo report to the Richards committee was blocked by Healy, not Hadley, one reading of an ambiguous sentence. □

Flare path

SIR — In describing the Japanese Solar-A mission due for launch on 26 August 1991, to study solar flares (*Nature* 352, 96; 1991), David Swinbanks omitted to mention the substantial involvement of two British groups, University College London's Mullard Space Science Laboratory and the Space Science Department of the Rutherford Appleton Laboratory. These groups, together with US colleagues, are providing the Bragg crystal spectrometer to study the high-temperature plasma produced during solar flares.

J. L. CULHANE

Mullard Space Science Laboratory,
University College London,
Holmbury St Mary,
Dorking, Surrey RH5 6NT, UK

A. M. CRUISE
J. LANG

Space Science Department,
Rutherford Appleton Laboratory,
Chilton, Didcot,
Oxfordshire OX11 0QX, UK

The invisible-obstacle race

Joan Mason

In some respects, things are getting worse, not better, for women in science. Positive measures need to be taken for progress towards genuine equality of opportunity.

DARWIN was an enlightened man, a supporter of education for women; but he concluded that an evolutionary timescale was required for woman's intellectual powers to reach those of man, since feminine faculties were depressed by women's subjection across the millennia, while masculine faculties were enhanced by competition¹. "It must take many centuries for heredity to produce the missing five ounces of the female brain", as his friend George Romanes put it². The female, in brainpower, came between man and the child, the female of the 'civilized races' between males of the civilized and the savage races. This was the traditional view, going back to Aristotle and Hippocrates, sanctified by religion, institutionalized in society.

We are more hopeful nowadays. Women's presumed inferiority was self-fulfilling, as inferior opportunities engendered inferior achievement. We now know that cognitive differences between the sexes (while still in dispute) are small: most females resemble most males. The historical shadow, however, is longer than we think. The position of women in science, in particular, is in some respects worsening, in others improving painfully slowly.

Decline and segregation

Sincere efforts, indeed, are being made towards equal opportunities for women. But coeducation is less good for girls and women than for boys and men. Fewer girls opt for physical science in mixed than in single-sex schools. When women's colleges go mixed, senior positions are soon filled by men, but if they do not they lose students and staff to mixed colleges with better resources. When men's colleges go mixed the familiar pyramidal distribution emerges. At Cambridge University, the former men's colleges (which began to admit women in 1972) now have an average 34% of women undergraduates, but less than 7% of women fellows.

Vertical segregation of women, in the language of the labour market, is marked in science as in other professions. Only 3% of professors and heads of department are women in UK universities³ (6.7% in polytechnics); comparative figures for women of professional rank are 5% in West Germany, 9% in France and 10% in the United States⁴. These numbers, little changed

over the past decade, conceal the horizontal segregation: women teach the (so-called) softer subjects, and receive lower pay than men doing similar work⁴.

Women chemists in the United Kingdom, for example, now comprise 35% of undergraduates, 24% of graduate students, 22% of postdoctoral researchers, 5.5% of lecturers, 1.5% of senior lecturers, 1% of readers and 0% of professors. The proportion of senior women is slightly higher in the United States: 5% of chemistry faculty with professional rank at PhD-granting institutions, and 16% in BS-only institutions. In the Royal Society of Chemistry⁵, women make up 37.5% of student members, 25% of graduate members, 8.5% of members, but only 1.7% of fellows; the

proportions are similar in physics and mathematics⁶⁻⁸. The Association for Women in Mathematics in the United States reports that women now earn about half of the bachelors' degrees and nearly a quarter of the doctoral degrees, but in the mathematics departments of Berkeley, Chicago, Harvard, MIT, Princeton and Yale combined, there is only one tenured woman⁹. Even in biology, the numbers of senior women are low: only 11% of members of the UK Physiology Society are women¹⁰.

Room at the top

Much of the best talent in science in the United Kingdom is, or should be, within the Royal Society. Forty-six years since women were admitted, the society contained 32 (2.9%) women fellows (excluding special categories) in December 1990¹¹. Figure 1 shows that even this small proportion has fallen systematically from its all-time high of 3.35% in 1979. Figure 2 shows that the proportion of women elected has not kept pace with increase in the number of fellows, a decline, some say, that may well continue. Over the past decade, 3% of the newly elected fellows have been women. (The society has just elected its first woman officer, the embryologist Anne McLaren.)

The Royal Society is not, of course, unique in its composition. National academies of science contain only 2-4% of women, except for those of Japan (all-male) and the Soviet Union (1%). The Australian Academy of Sciences has 2.3% women, the equivalents in India 2.4%, Sweden 3.1%, Canada 3.2%, China 3.75% (in 1981)¹², Denmark 4% and France 3.2% (having voted down Marie Curie in 1911, the year of her second Nobel prize; their first woman member was admitted in 1967). The US National Academy of Sciences has 4.1% women, and that proportion is increasing: over the past 6 years, 10.6% of new members have been women.

Horizontal segregation is also evident. Of the 47 women elected to the Royal Society since 1945, 39 (83%) have been biologists; among the physical scientists are no mainstream chemists or physicists, or engineers. Of the 80 women elected to the US National Academy of Sciences since 1925, 83% (again) have been biologists. I am reminded of Margaret Rossiter's categorization¹³ of

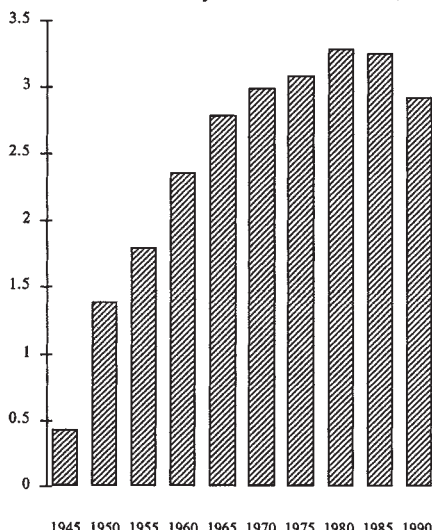


FIG. 1 Women as a percentage of fellows of the Royal Society.

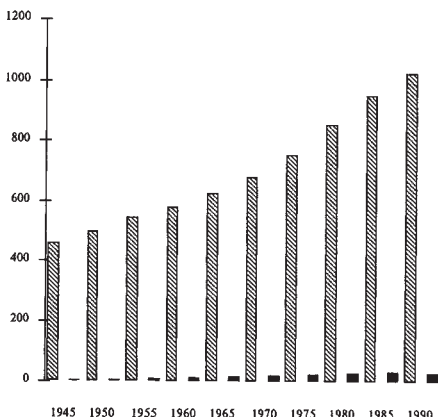


FIG. 2 Numbers of male (hatched boxes) and female (black boxes) fellows of the Royal Society.

sciences in which women's entry was facilitated early on in subjects like astronomy and crystallography by the need for tedious, low-paid data-collection and processing.

Missing women

The problem for women nowadays is less of access and more of equal opportunities for advancement and of keeping going once started. But the story of women's marking time, setback or drop-out as they marry or have children, has to be pieced together from the holes in the statistics of the survivors.

Women engineers were first to organize, when women's lower pay was seen as unfair competition (a double handicap). The UK Women's Engineering Society was formed in 1919 after women engineers who had taken the place of men during the war were sacked. The story in the United States is similar, the Society of Women Engineers taking shape after the Second World War. Women are now 15% of new engineering graduates in the United States¹⁴ (about 10% in the United Kingdom) and 6–7% of professional engineers (5% in the United Kingdom).

Learned societies have been looking into these problems^{5–8}. It seems that about a half of women drop out of science and engineering not long after they qualify. A 1988 survey of the Royal Society of Chemistry⁵ found that 70% of the women respondents had no children (compared with 34% of the men); only 6% had self-supporting children compared with 28% of men. Women chemists' average earnings were less than men's at all ages, falling gradually to about 75% of men's, reflecting loss of status while raising children. This is a general pattern.

Invisible obstacles

Beginning in the nursery is the received wisdom in which girls are expected to help and to care, boys to compete (and to choose); this leads to more girls in nursing than in medicine, or in biology than in physics or mathematics. Particularly sad is the recent history of computing. The proportion of girls taking courses at school or college has dropped steadily over the decade, as computing (other than word-processing) came to be seen as a masculine field, allied to mathematics and technology.

Women, of course, have their children at an age when scientists are establishing their careers. But it is not so much bearing children, as raising them, that takes the time. Institutions provide back-up systems, including technical and secretarial support, for the efficient working of their male executives; home-making support is supplied by wives. For women executives, in some of their most pro-

ductive years, essential support could include near-workplace child-care facilities, or domestic help. But such provision falls far short of the demand. Employers commonly ask a potential female employee what *her* child-care arrangements are; employers offer few such facilities because they are expensive. The United Kingdom is bottom of the European Economic Community league tables for paid maternity leave, job security and publicly funded child-care services for children under school age and for primary school children outside school hours¹⁵. The cost of child care is not even tax-deductible.

Traditionally, the (male) 'ecosystem' is designed for people who, once on the bottom rung, climb the ladder steadily, keeping up with, or ahead of, the competition. Many women take a job with lower status when their husbands move to a better job, or cope with 'telegamy' (a word for couples who commute long distances to spend weekends together). With essential family support, with luck, some women might maintain a (nearly) 40-hour week for their professional work (compared with 69–70 hours a week for male high-achievers). Part-time working could help many women over a crucial period.

Women in the United Kingdom can find their way back into science through the Open University's Women in Technology programme^{16,17}. Daphne Jackson's scheme to help qualified women returners by raising contributions from sponsors should surely be reinstated on a permanent basis.

Affirmative action for the training and employment of women and minorities in science and engineering has been most vigorously pursued in the United States¹⁴ (including quotas imposed on companies in receipt of government contracts). Conventional universities train and re-train mature students, a trend that is spreading elsewhere. The experience of the Scandinavian countries, which lead in child-care provision and a better balance of men's and women's responsibilities in the home, has shown that positive steps are needed to help women make progress in traditionally male preserves.

The Royal Society of Canada¹⁸ is grasping the nettle by encouraging members to nominate qualified women; the proportion of new women members in the arts and sciences has risen from 4 to 14% during the 1980s. The programme aims to "improve the climate for women in Canadian society as a whole . . . A fundamental principle of equity of opportunity is that the criteria and standards of excellence by which nominees are judged remain intact . . . A visible sign of intent would be the election of women members to its Executives, Councils and Committees"¹⁷.

A pioneer in the United Kingdom is University College London. Its policy of equal opportunities (including attention to problems of family and child care) has achieved a proportion of women professors, 9%, which is three times the national average; achieved, according to the former provost, Sir James Lighthill, by "a simple refusal to be blinkered by traditional preconceptions and misconceptions". King's College Cambridge, which went mixed in 1972, has adopted a similar policy in the face of a 530-year masculine tradition¹⁹. But to describe these efforts, admirable though they are, is to illustrate their rarity.

Girls, it is understood, need role models; but men and women, also, need exemplars. Based on women's relatively poor showing, past and present, it is a common, tacit belief that women are less good at science (arts, management, politics . . .) than men, and so the cycle repeats. Getting married and having children is a prerogative of both sexes: traditionally, the children and the home have come first for the wife, the career for the husband. What does one say to a young woman scientist who asks when, from the point of view of her career, is the best time to have children? When will equal-opportunities provision receive proper funding; and when will the legislation receive a proper set of teeth? □

Joan Mason is attached to the Department of Chemistry of the Open University, and to the Department of the History and Philosophy of Science, Free School Lane, Cambridge CB2 3RH, UK.

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
2. Romanes, G. J. *Nineteenth Century* **21**, 654 (1887).
3. *Women at the Top* (The Hansard Society for Parliamentary Government, London, 1990).
4. Sivard, R. L. *Women. A World Survey* (World Priorities, Box 25140, Washington, DC, 1985).
5. *Education, Employment and Attitudes of Men and Women Members* (Royal Society of Chemistry, London, 1989).
6. *Girls and Physics* (The Royal Society, London, 1982).
7. *Survey of Women Members* (Institute of Physics, London, 1986).
8. *Girls and Mathematics* (The Royal Society, London, 1984).
9. *Science* **251**, 628 (1991).
10. Bindman, L. J., Brading, A. F. & Tansey, E. M. *J. Physiol. Lond.* **430**, 101P (1990).
11. *Year Books of the Royal Society of London* (Royal Society of London).

12. Yao Shuping *XVIIth Congress of the International Union for the History and Philosophy of Science* (Berkeley, California, 1985).
13. Rossiter, M. W. *Women Scientists in America* (Johns Hopkins, Baltimore, 1982).
14. *Women and Minorities in Science and Engineering* (National Science Foundation, Washington, DC, 1982 et seq.).
15. *Childcare in the European Communities 1985–1990* (Commission of the European Communities, Brussels, 1990).
16. Carter, R. & Kirkup, G. *Women in Engineering: A Good Place to Be?* (Macmillan, London, 1990).
17. McRae, S., Devine, F. & Lakey, J. *Women into Engineering and Science: Employers' Policies and Practices* (Policy Studies Institute, London, 1991).
18. *Plan for the Advancement of Women in Scholarship* (Royal Society of Canada, Ottawa, 1989).
19. Spurling, A. *Report of the Women in Higher Education 1988–90* (King's College, Cambridge, 1990).

Can any anti-particle make a probe?

Positron probes (of the solid state) are already a booming business and may soon be applied to the study of liquids. Can anti-proton probes be far away?

EACH new particle of matter discovered may most immediately be of interest to those with theories to check (or invent, as the case may be), but one thing is certain: somebody will eventually use the particle as a probe for some purpose or another. And that is likely to be true even if the newly discovered particle is inherently unstable, or if it interacts with and is destroyed by some component of the environment in which it finds itself. What matters is merely that the particle should last for longer than the timescale of the phenomena studied. And the instability may then be an advantage, making plain either where the particle was before it ceased to exist or some other attribute of its state of being.

This generalization is best justified by the growth, in the past three decades, of the use of the positron as a probe, mostly of the solid state. It is an improbable development. Although positrons are inherently stable, with almost as much right as ordinary negatively charged electrons to continue indefinitely to exist, the whole world knows that positrons readily annihilate with electrons, so that their lifetime in the real world is best measured in, say, picoseconds (10^{-12} seconds). By way of compensation, the mutual annihilation of a positron and an electron yields two photons with a characteristic energy that peaks at 511 keV — next to the 21-cm line of hydrogen, the radiation frequency at which cosmologically inclined astrophysicists are always on the look-out, in observations of the galactic centre, for example.

Hitherto, the solid state has been the chief beneficiary. The principles are quite simple, at least on Page One. The yardstick of stability for a positron in the real world is the lifetime of positronium — the hydrogen 'atom' in which a positron has replaced the proton. Naturally, the singlet state (in which the spins are anti-parallel) is the least stable, decaying in 125 picoseconds. Any intrinsically more rapid phenomenon is in principle observable.

Naturally, there are complications, practical and of interpretation. Making nearly mono-energetic beams of positrons whose energy compares with those characteristic of the processes in solids is not child's play; the favoured techniques do not rely on high-energy physics, but on artificially produced isotopes such as ^{22}Na or ^{58}Ni , which are positron emitters. But, as with β -emitters, these materials yield particles with a Fermi spectrum ranging

from zero to some maximum.

The trick of making a mono-energetic beam of reasonable but still low energy usually entails some interaction with a solid within which the energy of incident positrons is first reduced to the average thermal energy, whereupon some proportion of the incident particles is afterwards re-emitted by scattering, but with an energy that is a measure of the work function — the potential difference between the interior and the exterior of the solid. Then, usually, there must follow devices for accelerating and shaping the beam. All that is intricate stuff.

And there is worse to come. Positrons fired into a solid are usually thermalized within a few picoseconds, whereafter several things can happen. The endpoint of a positron's existence is the formation of positronium, but until that happens it can diffuse. And what, then, is measured? The standard diagnostic appears to be the Doppler shift from the frequency of 511-keV caused by the momentum of the positron, itself dominated (in metals, at least) by the energy of electrons in the conduction band. But since there are many incident positrons, the trick is to measure the shape of a photon-emission band centred on 511 keV and then to make sense of it.

Even so, positrons not immediately captured are likely to be trapped at simple ionic vacancies in the crystal lattice of a metal. This process is efficient because the quantum wavelength of a positron thermalized to, say, room temperature is still many times the typical lattice spacing in solids. And, once trapped in a vacancy, the wave function of the positron will collapse geometrically as it substitutes for the missing ion, the chances of its forming positronium will decrease and, when it does link up with an electron from the solid, the momenta of both particles will be so small that two gamma rays eventually emitted in the annihilation will almost exactly compensate in momentum, so that their Doppler shift will be only small.

That explains the interest of positrons as probes in the exploration of defects in solids and of the detailed structure of surfaces. Different levels in a solid can be probed by different choices of incident energy. But it is only right to acknowledge that the hardy band of men and women working in this field are among those who most valiantly extract meaning from fuzzy data. Each of them deserves some kind of medal.

So what is new? Two things, both in the same issue of *Physical Review Letters* (2 September). First, E. Gramsch, K.G. Lynn, J. Throwe and I. Kanazawa, from the Brookhaven National Laboratory, have extended the use of positrons as probes from solids to liquids, and with surprising results (67, 1,282; 1991). The incentive has been to see whether positron probes can throw light on the structure of liquids, for which purpose they have studied positron annihilation in gallium and bismuth at the melting transition.

One expectation is readily confirmed. At the transition from solid to liquid, the diffusion length of positrons in either Ga or Bi is abruptly decreased (by a factor 10 in the former and 3 in the latter). Simplemindedly, this means that, in a liquid, there are fewer special sites at which positrons may hide from contact with electrons. The surprise is that, in molten gallium, the diffusion length appears to increase exponentially with increasing temperature, and not linearly as in bismuth.

The explanation offered is that positrons in liquid gallium are trapped in temporary potential wells formed by fluctuations of ionic structure and then, when that disappears, hop to the next nearest such fluctuation. What matters is that the wavelength of the positron is still quite long, but the picture cries out for a little calculation.

And what else? If the charge-counterpart of the electron can be a probe, why not also the charge-counterpart of the proton, the anti-proton? It is too soon to say that that has been done, but a Japanese group seems to be on the way. M. Iwasaki *et al.* from the University of Tokyo used a beam of anti-protons from the TRISTRAN accelerator at the National Laboratory for High-Energy Physics at Tsukuba to look for long lived anti-protons trapped in liquid helium (67, 1,246; 1991).

The outcome has been well-flagged, both theoretically and in experiments with negative pions. In principle, it should be possible for an anti-proton to replace an electron in a helium atom, when its lifetime should be much longer than the usual few picoseconds. What the Japanese find is that the annihilation of 3.6 per cent of the particles stopped in liquid helium is delayed, in some cases for as long as 30 microseconds — a factor of a million or so. For the time being, the trick seems to work only with helium, but it can only be a matter of time before somebody makes a probe of it.

John Maddox

Minute particulars

Robert W. Cahn

He who would do good to another must do it in Minute Particulars. General Good is the plea of the scoundrel, hypocrite and flatterer; for Art and Science cannot exist but in minutely organized Particulars.

THUS William Blake, who depicted Newton (for him a man akin to a devil), brooding evilly over a pair of dividers.

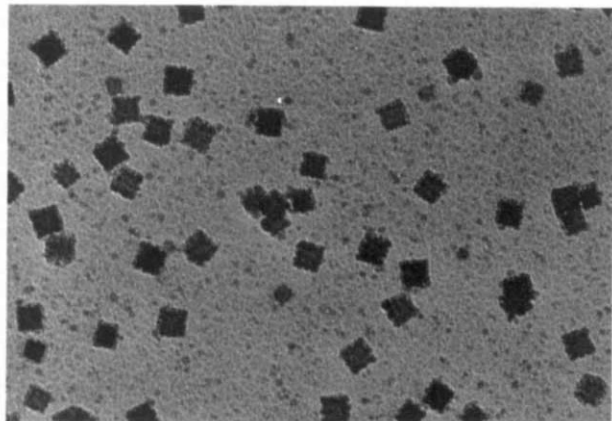


FIG. 1 Bright-field transmission electron micrograph of Mo deposit using a sputtering power of 150 W at 150 mTorr argon pressure (from ref. 2).

Nowadays, more than dividers are needed to identify minutely organized particulars. Ever since Leuwenhoek and Hooke, the organization of minute particulars has been examined down the tube of a microscope, at ever increasing levels of resolution. This reflection is prompted by a devilishly intriguing observation of *Minute Particulars*, published in *Science*¹. Edelstein and his collaborators have found evidence that tiny metal crystallites formed in a vapour assemble themselves on a substrate surface into larger blocks.

The groups at the Naval Research Laboratory in Washington, Bellcore and Johns Hopkins University have for some time been studying^{2,3} nanosized clusters, or crystallites, of molybdenum atoms, and 'cluster-assembled' materials, as the jargon goes, based on these tiny crystallites. They found that when they sputter molybdenum in argon at very low pressures, no distinct crystallites are formed but only a continuous deposit, coating an amorphous substrate. But at higher sputtering pressures, a population of very fine crystallites is made. They postulated that the very small mean free path characteristic of high gas pressure allows metal atoms to assemble into crystallites large enough to be stable and then to deposit as ready-formed crystallites on the carbon substrate.

Up to 100 mTorr (about 10^{-4} atmos-

pheres) of argon pressure, a continuous polycrystalline molybdenum film is formed; at or above 200 mTorr, a wide distribution of particle sizes results. Around 150 mTorr, however, a bimodal population of square particles results (Fig. 1); the smaller particles, scarcely resolved in the figure, average a cube edge of 4.8 nm (15 unit cell lengths), while the larger average 17.5 nm, both with a small mean deviation. There are no crystallites with edge lengths between 9 and 14 nm. The particles are undoubtedly body-centred cubic molybdenum, with a lattice parameter (measured by electron diffraction) within 1 per cent of that of bulk molybdenum, and the larger particles, at least, consist of cubes with their crystallographic (100) faces parallel to the substrate; the evidence, not conclusive, suggests that the smaller cubes are similarly arranged.

The authors conclude from a detailed examination of the morphology of the larger cubes that they were formed by self-assembly of the smaller cubes, three per edge of the larger cubes, so that the larger cubes are each made up of 27 smaller cubes. The imperfect joins of the small cubes can just be made out in Fig. 1. The fact that the average edge length of the larger cubes is a non-integral multiple, 3.6 times, of that of the smaller cubes, is not inconsistent with this model as the packing of the smaller cubes appears to be imperfect. The authors point out that puzzles remain, such as the reason for the cubic shape of the smaller particles before they assemble.

The self-assembly observed here appears to be a case of coalescence on the substrate. This kind of process has often been observed in studies of the growth of thin films, although such regular assembly of ready-made cubes is new, and in the older work, the original nucleation of crystallites took place on the substrate and not in flight. The geometry and mechanisms of such coalescence have recently been reviewed by Pashley⁴. The process is liquid-like (Fig. 2): even if the particles have crystallographic contours before and after coalescing, they do not while the process is in progress. It is hard to see how the process of Fig. 2, a kind of two-dimensional Ostwald ripening, could

generate the morphology of Fig. 1. Moreover, the process discussed by Pashley involves epitaxial deposits, constrained in orientation by the substrate, whereas in the molybdenum work the substrate is amorphous and (as can be seen from the lack of preferred orientation in Fig. 1) there is no epitaxy.

The process of self-assembly of nanocubes, starting from a population of cubes lying flat but oriented in random azimuths, would entail rotation of particles as they move across the surface. This kind of random migration, followed by rotation and coalescence, was observed long ago by Masson and Kern, who were examining metal deposition onto rocksalt; in this situation, there is epitaxy. I reported at the time⁵ on the observations and their vigorous interpretation by a number of commentators. Most of the discussions — for instance, Reiss's treatment of the orientation dependence of tiny nuclei on a substrate⁶ — really referred to a situation of epitaxy and their relevance to the new work with an amorphous substrate is uncertain. It would be interesting to see the experiment by Edelstein *et al.* repeated with a crystalline, epitaxy-inducing substrate, on which the orientation of deposited ready-made crystallites might be constrained.

Other possible developments present themselves. Thus in 1955, M. M. Nicolson (an able physicist who unfortunately died very young) calculated from first principles⁷ how the lattice parameter of tiny magnesium oxide crystals should change with crystallite size, because of the proportionately larger influence of surface tension: for MgO, the predicted decrease of lattice parameter was about 0.5 per cent for an edge length of 10 nm. Nicolson's experiments showed a decrease somewhat less than half the predicted amount, for both MgO and NaCl. If it proves possible to make precision measurements of lattice parameter on the assembled molybdenum crystals, using X-rays rather than electrons (and if theory predicts an effect similar to that in ionic salts, which is not a foregone

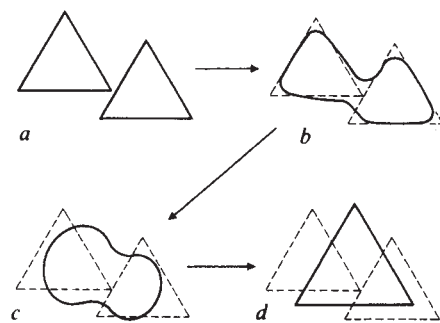


FIG. 2 The sequence of shape changes during the liquid-like coalescence of two triangular epitaxial nuclei on a crystalline substrate (from ref. 4).

conclusion), it would be interesting to see whether the lattice parameter change is that characteristic of a 4.8 nm crystal or of a 17.5 nm crystal: in other words, does the assembled crystal really have the properties corresponding to the assembled size? In this connection, the team which produced the molybdenum nanocrystals has recently been able to disassemble the 17.5 nm crystals by manipulation with the probe of a scanning tunnelling microscope (A. S. Edelstein, personal communication), indicating that the assembly is far from perfect. The nature and strength of the forces holding the crystallites together remain to be determined.

Also, micrographic or microcalorimetric experiments to measure the melting temperatures of the assembled molybdenum crystals might give a similar insight, as Kofman *et al.*^{8,9} showed that lead nanocrystals embedded in SiO melt at temperatures well below the bulk temperatures depending on the crystallite size (in the latest of a succession of experiments demonstrating premature

melting of very small metal crystallites of various kinds). Another interesting approach would be to examine crystals like those in Fig. 1 by high-resolution electron microscopy, in association with computer simulations of the images as recently demonstrated by Kirkland *et al.*¹⁰. The *General Good* might be enhanced by such experiments. □

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

1. Edelstein, A. S., Chow G. M., Altman, E. I., Colton, R. J. & Hwang, D. M. *Science* **251**, 1590–1592 (1991).
2. Chow, G. M., Chien, C. L. & Edelstein, A. S. *J. Mater. Res.* **6**, 8–17 (1991).
3. Chow, G. M. *et al.* *J. Mater. Res.* **6**, 737–743 (1991).
4. Pashley, D. W. in *Processing of Metals and Alloys* (ed. Cahn, R. W.) 290–328 (VCH, Weinheim, 1991).
5. Cahn, R. W. *Nature phys. Sci.* **241**, 1 (1973); *Nature* **246**, 66–67 (1973).
6. Reiss, H. *J. appl. Phys.* **39**, 5046 (1968).
7. Nicolson, M. M. *Proc. R. Soc.* **228**, 490–510 (1955).
8. Garrigos, R., Cheyssac, P. & Kofman, R. *Z. Phys.* **12**, 497–500 (1989).
9. Kofman, R., Cheyssac, P., Garrigos, R., Lereah, Y. & Deutscher, G. *Z. Phys.* **20**, 267–271 (1991).
10. Kirkland, A. I., Jefferson, D. A., Tang, D. & Edwards, P. P. *Proc. R. Soc.* **A434**, 279–296 (1991).

COGNITIVE NEUROPSYCHOLOGY

Writing without vowels

John C. Marshall

In this issue (page 258¹) Roberto Cubelli reports a novel (and quite unexpected) form of writing deficit. Two Italian-speaking patients who had suffered damage to the left hemispheres of their brains were selectively impaired on writing vowels.

Disorders of writing and spelling (the dysgraphias) are a frequent consequence of damage to the left hemisphere. If adults who were previously fully literate, suffer stroke, tumour or trauma that involves left parieto-occipital cortex in particular, it is highly likely that significant reading and writing impairments will result. These problems will usually, but not inevitably, be found in the context of more widespread disorder in the expression and comprehension of spoken language (aphasia), with concomitant involvement of temporal cortex.

The qualitative nature of acquired dysgraphia can assume many forms. Some patients write what they hear, and will thus (in English) transcribe yacht as yot, laugh as laf, or room as rewm². Equivalent errors from French patients include photo written as fauto, and pigeon as pijon³. Mistakes of this type occur both in spontaneous writing and in writing to dictation. Other patients write (approximately) what they (or the examiners) mean. Cake may be written as bun, star as moon, or nephew as uncle⁴. Comparable errors reported in French are crabe written as homard, auto as voiture, and

biche as gazelle⁵.

Some patients show that they recognize their own spelling problem by simply leaving out the letters they are unsure of. For instance, one patient attempting to produce the names of countries wrote: *In a*; *Tur y*; and *C na*⁶. Cubelli's two subjects, who had suffered left-hemisphere infarcts, manifested a curious variant of this deficit.

The first (C. F.) was similar to the case just mentioned in that he left spaces for the 'unavailable' letters; but unlike that earlier case, the missing items were always vowels. For example, requested to write the name of the town in which he lived, C. F. wrote B L G N. There is no obvious reason why vowels should be more difficult to 'remember' than consonants or more difficult to execute motorically. One might, however, reflect that when y wrt wtht vwls the message is still (fairly) clear, but e ou ie iou oe The former option was, of course, adopted for Phoenician script, and can still be seen (to a first approximation) in modern (unpointed) Hebrew⁷. C. F.'s stroke had resulted in a right-sided hemiplegia, and the patient was therefore forced to write with his left (non-dominant) hand. It is thus possible (but intuitively unlikely) that C. F.'s pattern of performance was a 'strategic' adaptation to the difficulty of writing with his left hand; Fig. 1 (page 259) of Cubelli's report shows some shakiness in the ex-

ecution of the (upper case) letters that C. F. produced.

The second case (C. W.), however, had no hemiplegia and wrote with his (preferred) right hand. His performance cannot by any stretch of the imagination be made to fit this 'strategic' hypothesis. C. W. did make some errors on consonants, but there was an overwhelming preponderance of vowel errors. More crucially, C. W. did not (typically) omit vowels (as C. F. did); rather, 90 per cent of his errors were substitutions or transpositions. For example, dietro (behind) was written as diatro, and caro (dear) as cora. These two responses, like the vast majority that Cubelli reports C. W. as making, are possible but not actual Italian words. The pattern of performance was constant across a range of relevant tasks: written and oral spelling, typing, and delayed copying. In all instances, vowel errors were significantly more frequent than consonant errors.

The critical question is where in the functional architecture of the spelling system could such errors arise. C. W.'s error rate is not affected by lexical variables: he makes the same proportion of errors irrespective of whether he is writing high or low frequency words, abstract or concrete words, or indeed words or nonwords. Likewise, the errors themselves are not constrained to be (attested) words. The disorder thus seems to have no involvement with central lexical processes.

Neither is it the case that C. W. has 'lost' the graphemic representations of vowels as such; he can write single vowels and syllables to dictation without error. Substitutions, transpositions and omissions occur only when a multisyllabic sequence of letters is required. The conclusion, then, is that there is impairment of a short-term memory system specific to written (and oral) spelling — the 'graphemic buffer'.

The deeper question of why this graphemic buffer should be sensitive to the phonologically based distinction between vowels and consonants remains open. C. W. has good (but not perfect) immediate repetition of the words and nonwords that give rise to such high levels of writing errors (Table 1, page 259). One would like to know if delayed oral repetition showed errors selectively concentrated on vowels. Cubelli argues that "the consonant/vowel status of graphemes is differentially specified in

1. Cubelli, R. *Nature* **353**, 258–260 (1991).
2. Newcombe, F. & Marshall, J. C. in *Surface Dyslexia* (eds Patterson, K. E. *et al.*) 35–51 (Lawrence Erlbaum Associates, London, 1985).
3. Beauvois, M. F. & Dérouesné, J. *Brain* **104**, 21 (1981).
4. Marshall, J. C. & Newcombe, F. *Neuropsychologia* **4**, 169–176 (1966).
5. Michel, F. *Lyon Médical* **241**, 141–149 (1979).
6. Morton, J. in *Cognitive Processes in Spelling* (ed. Frith, U.) 117–133 (Academic, London, 1980).
7. Barr, J. in *Writing without Letters* (ed. Hass, W.) 71–100 (Manchester University Press, 1976).

the spelling process and may be selectively affected following brain damage" but this selectivity may not be truly specific to writing. It is furthermore difficult to imagine that the direction of the effect could be reversed in other patients. Roberto Cubelli has shown that a dysgraphic patient could write that

name as R B R T C B L L ; but the odds do not appear favourable that we shall ever find his name transcribed as O E O U E I. □

John C. Marshall is in the Neuropsychology Unit, University Department of Clinical Neurology, The Radcliffe Infirmary, Oxford OX2 6HE, UK.

CLIMATE CHANGE

Is water vapour understood?

R. L. Jones and J. F. B. Mitchell

ON page 244 of this issue¹, Kelly and colleagues describe a marked imbalance in the water to be found in winter in the upper tropospheres of the Northern and Southern Hemispheres. The observation may have considerable implications for modellers attempting to forecast the course of climate change.

Water vapour is the single most important greenhouse gas²⁻⁴. In the vapour phase in the upper troposphere it modulates the outgoing longwave radiation, and is an integral part of the formation of clouds, which affect both the incoming solar and outgoing longwave radiation fluxes⁵. Concentrations of water in the atmosphere are expected to change in response to changes in concentration of other greenhouse gases. For example, as increases in carbon dioxide warm the surface and atmosphere, more water evaporates from the surface and remains in the atmosphere. In fact the amount of water vapour which can be held increases exponentially with temperature. Because water vapour is such a strong greenhouse gas, this increase traps more longwave radiation, further warming the surface and troposphere and so amplifying the initial warming. This feedback is very significant and can increase the original perturbation by as much as a factor of two⁶ in the absence of other feedbacks.

It is thus vital for climate prediction models to be able to predict reliably changes in the global water vapour distribution. Hence an important test of any climate model is its capacity to reproduce the present distribution of water vapour and the extent to which it incorporates the processes which control water vapour in the atmosphere.

Global measurements of the total water content do exist⁷, and there are measurements in cloud-free conditions Equatorward of 60 degrees⁸. Kelly *et al.* now present new and highly accurate measurements of water vapour in the upper troposphere at latitudes poleward of 40 degrees. Although these are as yet only for winter months, they provide fresh information on the global distribution of water and, as the authors show, offer insights into the mechanisms which

control water concentrations in the upper troposphere.

In their paper, the authors present measurements of water vapour in the upper troposphere and low stratosphere of both hemispheres obtained from hygrometers on board a DC-8 and an ER-2 aircraft. The measurements themselves are striking in that they show that, at middle and high latitudes throughout the upper troposphere and lowest stratosphere, in winter there is two to four times more water vapour in the Northern Hemisphere than in the Southern. Kelly *et al.* then describe a case study which shows that dryness in the upper troposphere of the Southern Hemisphere at mid-latitudes can result from large-scale transport of that mid-latitude air to polar regions. There, adiabatic ascent occurs, cooling the air sufficiently to dehydrate it to the concentrations of water vapour observed at mid-latitudes. In essence, the authors' argument is that the observed interhemispheric water difference is a direct consequence of different polar temperatures in the upper troposphere of the two hemispheres in winter.

It is important to establish whether current climate models reproduce this asymmetric interhemispheric distribution of water (which incidentally does not appear so markedly in the relative humidity), and whether they incorporate the dehydration mechanism Kelly *et al.* describe. Together with G. Cookmartin, we have analysed results from a general circulation model (GCM)⁹, run for another purpose, to consider these questions. In the GCM simulations it was found that at high latitudes (polewards of 65 degrees) there is an asymmetry in the water vapour concentrations in the upper troposphere in the same sense as, but less marked (only up to a factor of two) than, that in the observations. At lower latitudes in the GCM, this asymmetry in the concentrations declines and even changes sign.

There are, of course, a number of difficulties with such a comparison. First, the model diagnostics used were seasonal and zonal (or ocean only) averages and were thus not strictly comparable to

the authors' data which had different geographical and temporal sampling. Because of this limited coverage there is also a question of the representativeness of the authors' data. For these reasons it is not yet possible to establish whether the discrepancies represent a fundamental limitation of the GCM. Clearly though, the data provide an important and stringent test of it.

The notion raised by Kelly *et al.* of dehydration of middle and upper tropospheric mid-latitude air in polar regions also raises some interesting questions. In GCMs, local water vapour content is controlled by a variety of processes, including condensation and evaporation, which depend strongly on local temperature, and transport of more or less moist air from elsewhere. It is important for the accurate simulation of the water vapour distribution, and the calculation of the water vapour feedback, that the balance of these controlling processes is correctly reproduced by the model. As Kelly *et al.* point out, there are indications that the flows in numerical models of the atmosphere tend to be too zonal in character¹⁰. If so, and if the contribution of Kelly *et al.*'s mechanism for dehydration is substantial, this balance may not be correct and, for example, GCMs will tend to restrict any water asymmetry to high latitudes, a feature of the results described above.

If, as Kelly *et al.* suggest, polar temperatures control water vapour concentrations at mid-latitudes, because mid-latitude and polar tropospheric temperatures respond similarly to climate forcing^{2,11} the simplistic conclusion is that current GCMs may still have the correct water vapour feedback in the extra-tropics, but for the wrong reason. The authors' main achievement, however, is to have demonstrated the value of accurate water vapour measurements in the upper troposphere for assessing climate models, and thereby show the need for a more systematic study. □

R. L. Jones is in the Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. J. F. B. Mitchell is in the Hadley Centre for Climate Prediction and Research, Meteorological Office, Bracknell, Berkshire RG12 2SZ, UK.

1. Kelly, K. K., Tuck, A. F. & Davies, T. *Nature* **353**, 244-247 (1991).
2. Houghton, J. T. *et al.* *Climate Change: The IPCC Scientific Assessment* (Cambridge University Press, 1990).
3. Raval, A. & Ramanathan, V. *Nature* **342**, 758-761 (1989).
4. Mitchell, J. F. B. *Rev. Geophys.* **27**, 115-139 (1990).
5. Ramanathan, V. *et al.* *Science* **243**, 57-63 (1989).
6. Manabe, S. & Wetherald, R. T. *J. Atmos. Sci.* **24**, 241-259 (1967).
7. Prabhakara, C. *et al.* *J. appl. Met.* **21**, 59-68 (1982).
8. Rind, D. *et al.* *Nature* **349**, 500-503 (1991).
9. Smith, R. N. B. *Q. J. R. Met. Soc.* **116**, 435-460 (1990).
10. Tuck, A. F. *et al.* *J. geophys. Res.* **94**, D9, 11181-11222 (1989).
11. Manabe, S. *et al.* *J. Climate* **4**, 785-818 (1991).

Flying in the face of reason

William J. Sutherland

BIRD migration is a puzzling phenomenon — why, for example, do some coastal birds traverse half the world whereas their conspecifics migrate only a fraction of this distance? New studies¹⁻⁹ have now determined the costs and benefits of different migration strategies by examining the same species at different latitudes. They show how an understanding of prey selection and energetics is necessary to explain why birds winter in tropical rather than temperate regions, and how the timing of migration depends upon the subtleties of the behaviour of the tropical prey.

Over the past two decades, Dutch ornithologists have carried out detailed observations on waders in the Wadden Sea which, with an annual peak count of two million waders, is the most important mudflat in Europe. Although some waders winter on the Wadden Sea, many of the birds present during spring and autumn spend the winter on the Banc d'Arguin, an area of tidal mudflats in Mauritania, West Africa. The Banc d'Arguin holds over two million waders, which is more than a third of the waders that use the estuaries of the East Atlantic. In some species different populations or subspecies winter at these very different latitudes¹. For several years the researchers from the Wadden Sea have also migrated south to study the consequences of wintering in the tropics compared with staying in temperate latitudes, and a recent volume of *Ardea** is devoted to their results.

The traditional explanation for migrating south is that the winter food supply is higher in the tropics. Early work¹⁰ showed however that, contrary to all expectations, there is a remarkably low density of prey at the Banc d'Arguin yet a high density of wading birds. The more detailed surveys² confirm that a low standing crop of food is present and that the bird densities are atypically high, almost four times greater than the average for temperate wintering areas. The estimated consumption of invertebrates by birds over the winter exceeds the standing crop.

The paradox of this imbalance can be resolved by considering the diet². Almost all bird species at the Banc d'Arguin feed on remarkably small prey; for example, for a range of wader species, the average mass of worms taken is

3–10 times lower than that of the smallest worms taken in Europe. As such small prey can have a much faster turnover, a low standing crop can sustain the high consumption. The striking exception to this pattern is the giant bloody



Caught out — a fiddler crab, away from its burrow, falls prey to a whimbrel on the Banc d'Arguin. (Jan van de Kam.)

cockle *Anadara senilis*, which is eaten by oystercatchers *Haematopus ostralegus*³: the fresh mass of a giant bloody cockle can equal that of an oystercatcher.

If waders at the Banc d'Arguin have to feed on small prey from a low standing crop, why should they fly so far to feed there? One of the main advantages of migrating south is that in the warmer climates less energy is expended in temperature regulation. From environmental measurements, the thermostatic costs for a knot *Caladris canutus* in West Africa were estimated at half the costs of a bird on the Wadden Sea¹. And studies¹¹ of sanderling *Calidris alba* in a range of sites show that the energy requirements vary markedly between latitudes. For instance, in Panama energy requirements were estimated at twice basal metabolic rate, but in New Jersey this value was about double again.

The daily energy expenditure of a range of captive birds in the Banc d'Arguin, measured using doubly labelled water⁴, indicated that, when corrected for thermoneutral conditions, the maintenance metabolic rate is on average 42 per cent higher for birds wintering in the Wadden Sea than for those wintering in the Banc d'Arguin. It seems moreover that birds do not adjust their expenditure according to local conditions. In some species, such as turnstone

Arenaria interpres, different populations winter at different latitudes. Three Mauritanian turnstone were flown to the Netherlands and kept with three captured during winter in the Wadden Sea⁴. The birds from the tropical wintering grounds maintained a lower daily energy expenditure (158 kJ d⁻¹) than did those from the temperate site (215 kJ d⁻¹), and the differences in metabolism were maintained in the captive birds when re-measured over a year later. Such persistent differences will restrict the ability of birds to change wintering grounds and are likely to have consequences for the range of habitats that they can use at other times of the year.

One effect of the low density of prey seems to be a reduction in the birds' ability to gain the fat and protein necessary to migrate north in the spring. A review of all the measurements of mass gain in waders demonstrates that the rate of gain is usually much lower in birds departing from tropical wintering areas than for the temperate sites⁵. And studies⁶⁻⁸ of whimbrels, *Numenius phaeopus*, show how their spring migration schedule is intimately geared to the annual cycle of their tropical prey species. The main prey of whimbrel in the Banc d'Arguin is the fiddler crab *Uca tangeri*, which

the birds usually capture either by dashing at the crabs that are feeding too far from their burrows or by standing motionless above a crab burrow waiting for the inhabitant to emerge. From late March onwards, during spring tides, the crabs leave their burrows to form herds at low water and, with no burrow to retreat into, they can easily be caught (see illustration).

In the spring the whimbrels must start gaining the stores of fat and protein necessary to fuel their migration to the European staging areas, but they cannot start increasing in weight until the switch in crab behaviour at the end of March. So the timing of fat accumulation is dictated by the prey⁶. The activity of the crabs is dependent upon the lunar cycle, and this work suggests that the time at which birds can start increasing in mass will vary from year to year. This is likely to affect the date at which they can depart. In accordance with this idea, the variation in date at which whimbrel arrive in the Netherlands was found to be positively correlated with the timing of the lunar cycle.

Once the fiddler crabs have formed herds the whimbrel are able to feed at a high rate, consuming 2–3 mg ash free dry mass per second over short periods. But the amount the birds can consume is constrained by the indigestible exo-

* This issue of *Ardea* is available for 55 Dutch florins from Stichting WIWO, U.v. Stuivenbergweg 4, 6644 AB Ewijk, The Netherlands.

skeletons of the crabs which act as a digestive bottleneck^{7,8}. As a result they have to pause for digestion and the maximum crude intake is then reduced to 1 mg s⁻¹, which means that they can only increase their body mass by 1.1 per cent per day. With this slow rate of mass accumulation they cannot gain sufficient reserves to fly to their European stop-over sites before the end of April.

Most of these wading birds breed in the Arctic and the work described in *Ardea* illustrates how the birds are dependent upon a combination of tropical, temperate and arctic regions. The populations of many species of waders breeding over a vast area of tundra can only persist because of the capacity of the Banc d'Arguin to sustain the high concentrations of wintering birds. Conversely, the studies show that the slow rate of fat accumulation in the Banc d'Arguin, combined with the need to replenish reserves again in temperate areas, means that this tropical wintering area is most suitable for birds which breed in the high Arctic where it is possible to breed only during a short, though intense, season¹².

Although many birds winter in the productive estuaries of Europe, these sites are most important as intermediate fattening grounds in which birds can replenish the reserves that enable them to continue their migration between Arctic breeding grounds and tropical wintering sites. The Wadden Sea and the estuaries of Britain are by far the most important European staging areas. Unfortunately much of the intertidal habitat of the European estuaries has already been destroyed and a large proportion of the remainder is threatened; of the internationally important estuaries in Britain, over 50 per cent face further habitat loss by direct land claim¹³. At the moment the Arctic breeding grounds and tropical wintering grounds seem relatively safe. But it remains to be seen whether there is the commitment to conserving the European estuaries. □

William J. Sutherland is in the School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

1. Piersma, T. et al. *Proc. Int. Ornithol. Congr.* Christchurch (in the press).
2. Zwarts, L. et al. *Ardea* **78**, 39–52 (1990).
3. Swennen, C. *Ardea* **78**, 53–62 (1990).
4. Klaassen, K. et al. *Ardea* **78**, 209–220 (1990).
5. Piersma, T. & Jukema, J. *Ardea* **78**, 315–338 (1990).
6. Zwarts, L. et al. *Ardea* **78**, 257–278 (1990).
7. Zwarts, L. *Ardea* **78**, 279–300 (1990).
8. Zwarts, L. & Blomert, A. M. *Ardea* **78**, 193–208 (1990).
9. Zwarts, L. et al. *Ardea* **78**, 339–364 (1990).
10. Engelman, M. et al. in *Coastal Waders and Wildfowl in Water* (eds Evans, P. R., Goss-Custard, J. D. & Hale, W. G.) 293–310 (Cambridge University Press, 1984).
11. Castro, G. thesis, Univ. Pennsylvania (1988).
12. Drent, R. & Piersma, T. in *Bird Migration* (ed. Gwinner, G.) 399–412 (Springer, Heidelberg, 1990).
13. Davidson, N. C. et al. *Nature Conservation and Estuaries in Great Britain* (Nature Conservancy Council, Peterborough, 1991).

The bird in the bush

John H. Ostrom

FIVE years ago, palaeontologists were astounded by news reports that Dr Sankar Chatterjee of Texas Technical University in Lubbock, Texas, had discovered the remains of an ancient fossil bird in the Dockum formation in Texas — one older by 75 million years than *Archaeopteryx*, itself around 150 million years old. If true, the discovery would extend the stratigraphic range of birds by half. *Archaeopteryx*, known from six skeletal specimens, has long been recognized as the oldest unambiguous fossil bird and is widely accepted as providing the most important evidence about the origin of birds and the enigma of the origin of avian flight. It seemed unthinkable that birds 75 million years more ancient could have existed.

Interest

The announcement provoked a huge amount of interest, but when it was not followed up promptly with any formal documentation, interest turned to puzzlement. Now, however, as described in *Nature* in June (351, 677–678; 1991) Chatterjee has delivered a paper in *Philosophical Transactions of the Royal Society* (B332, 277–342; 1991) to back up his claim. What are palaeontologists, and the world at large, to make of it?

Sad to say, for all its length, little support for the claim is to be found in the paper. It deals only with the cranial material of the specimen (which Chatterjee calls *Protoavis texensis*). This material is fragmentary, disarticulated and severely crushed, and therefore subject to different interpretations — and even different identifications. Every fragment has been extracted from the original matrix so that the original spatial associations are no longer available.

The reader is thus presented with an incomplete analysis and an inadequate sense of the poor quality of preservation. Even Chatterjee's promise of a sequel, in which he says he will deal with the postcranial remains and discuss the origin of avian flight, cannot remove the strong feelings that emerge after reading the report in detail — ones of disbelief about the claim and disappointment about the way it has been presented.

My disbelief stems from the author's failure to provide what I consider to be firm evidence that the specimen is indeed avian in nature; I have myself seen the material (but briefly, and in less than ideal circumstances) and consider that key fragments may well be beyond identification. Disappointment stems from the belatedness of the paper's appearance and its shortcomings.

What we have are photographs of minuscule size and marginal quality, arranged in a confusing format (on occasion with 20 or more objects to a page) with brief and sometimes cryptic captions that are discouraging to even the most dedicated readers. Considering the potential importance of the claim that the specimen is a bird from the Triassic, it is astonishing that there are no stereo photo pairs (to give three-dimensional images) provided in support. Instead the report is padded with accounts of features that are not without interest but are irrelevant to the main point, such as those under "Neurosensory organs" (the brain, the eye, the ear) and "Cranial kinesis and quadrate movement", with its sections on the streptostylic quadrate and upper jaw mobility. This rather than addressing the central questions. Where are the feathers, an avian coracoid or a keeled sternum? Where is the wishbone? What about the carpometacarpus — the fused wrist and hand bones so characteristic of all birds (except *Archaeopteryx*)? Moreover, statement of Triassic age requires more than the passing comments to be found under "Geologic setting". And strange comments such as "Other missing bones [sic] show some sort of *post mortem* disturbance" do not add credibility.

Most notable by its absence is consideration of what else the remains might be. Chatterjee describes the cranial fragments from an avian perspective only, without admitting the prospect that they might be squamatan, or crocodilian, or pterosaurian, or rhynchocephalian. There are many other possibilities.

Unease

So I am left with a deep sense of unease. It is understandable that Chatterjee has wished to make the most of his find, but it is surely time that others should now be able properly to evaluate his claims. It is said the specimen will be made available for examination, by 'experts', but only in Lubbock. To my mind that falls short of reasonable access. Enough is enough — the material should now be transferred for a reasonable period to the Natural History Museum laboratories at the Smithsonian Institution where it can be the subject of detailed and independent scrutiny. Only then will Chatterjee have met his obligation to the research community, for in this paper he has failed to do so. □

John H. Ostrom is in the Peabody Museum of Natural History, Yale University, PO Box 6666, New Haven, Connecticut 06511, USA.

A vanishing interface

Philip Ball

THE distinction between a continuum and a molecular description is what, in large part, divides dynamical and statistical-mechanical (equilibrium) theories of fluid behaviour. A recent meeting* bore the good news that in surface and colloid science, where the molecular-scale details can seldom be ignored, this divide between statics and dynamics is narrowing. For chemical engineers a dynamical viewpoint has always seemed natural, of course: their concern has long been with problems of flow, rheology, dynamical instabilities. But it is reassuring that the microscopic theories now being brought to bear on these issues are, on the whole, equal to the challenge.

An obstacle to fundamental (as distinct from empirical) understanding of surface dynamics is the difficulty of realizing experimentally the idealized conditions that the theories must take as their starting point. The vagaries of a liquid droplet spreading on a surface (such as hysteresis, for advancing and receding droplets, in the angle of contact between the air-liquid and solid-liquid interfaces, or pinning of the 'contact line' at the film's edge to surface imperfections) are a useful measure of surface characteristics for engineers, but the bane of physicists trying to understand how spreading occurs.

For a liquid spreading on perfectly smooth, uniform surfaces, a hydrodynamical analysis predicts proportionality between the cube of the contact angle θ and the velocity of the advancing contact line. But real surfaces are seldom so ideal: J. F. Joanny (Institut Charles Sadron, Strasbourg) described how this spreading behaviour should be modified by surface defects. A contact line that becomes snagged on these is somewhat like a string under tension, but softer. A string relaxes harmonically when deformed and released, with a relaxation time τ proportional to the inverse square of the wavevector of deformation q . For the contact line, on the other hand, deformation affects also the two-dimensional liquid-air interface, with the result that surface point defects induce a smooth deformation and τ varies with the inverse of q .

T. Ondarçuhu and M. Veyssié (Collège de France) have now devised a way to test this prediction experimentally¹. They generate a homogeneous surface with a specified surface energy by coating a silicone wafer with long-chain aliphatic or fluorinated silanes, and

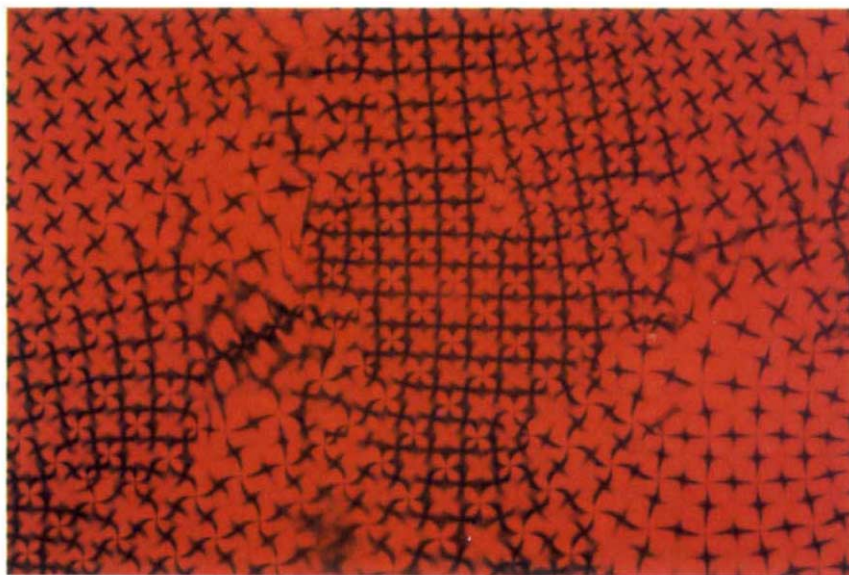
introduce a deformation of a single wavelength by merging a linear liquid front (silicone oils are used) with regularly spaced circular droplets. When gravitational effects are not important (that is, when q is not too small), the theory is vindicated: an encouraging basis on which an understanding of real-world spreading can develop.

The driving force behind droplet spreading on a horizontal surface (when the film thickness is not so large that gravity plays a part) is the lowering of surface energy, which is decreased when the substrate is either partly or totally covered by the liquid film (respectively partial or complete wetting). In the former case, a liquid film spread forcibly over the entire surface will retract to the equilibrium configuration in which parts of the surface are exposed again. This

'dewetting' process, relevant for example to the problem of keeping windscreens free from obscuring liquid films, also obeys the θ^3 dynamics² (C. Redon, Université Pierre et Marie Curie).

If the spreading rate is increased by driving the flow using, for example, gravitational³ or centrifugal⁴ forces, a linear contact line becomes unstable even in the absence of surface heterogeneity. A fingering instability with a well-defined wavelength then develops spontaneously³, as seen in the 'tears of wine' that form in a partly filled glass (A. M. Cazabat, Collège de France). The Marangoni effect — flow induced by gradients in surface tension — can also produce an instability. Surface-tension gradients can be created, for example, by imposing a temperature gradient on the flow⁵ or by the presence of surfactants (S. Troian, Exxon).

Like films overlying a solid surface, 'free-standing' soap films, such as those that constitute bubble walls, attain a particular equilibrium thickness, but the



In the ternary mixture of water, oil and a surfactant, physicists interested in phase transitions have tapped an extraordinarily rich vein. The phase diagram abounds with phenomena inviting study: percolation, self-assembly, wetting and criticality to name a few. A surprisingly simple lattice model of the three-component system, proposed by Widom¹¹ and now extended by K. Dawson (Berkeley), can reproduce much of this behaviour, such as the percolation transition from oil-filled micelles to a bicontinuous oil/water phase interfaced by surfactant. To minimize the surface energy of the amphiphilic layer, the system can adopt periodic structures described by 'minimal surfaces' (S. T. Hyde, Australian National University). The classification and stability of these structures can be understood by borrow-

ing concepts from topology.

Elsewhere in the phase diagram, surfactants may self-assemble into vesicles, lamellar phases or cylindrical micelles. These can be studied by rapid quenching of the solution, freezing the surfactant structures in a vitrified matrix which can then be probed by electron microscopy (Y. Talmon, Technion). An intriguing variant of the lamellar phase is obtained by incorporating small magnetic particles (around 10 nm in size), stabilized by coating them with amphiphiles (C. Quillet and P. Fabré, Collège de France). The lamellae of this 'ferrosmectic' phase can be realigned by an applied magnetic field. The figure shows a network of focal conic defects in the ferrosmectic phase viewed through a polarizing microscope. □

* 7th International Conference on Surface and Colloid Science Compiègne, France, 7–13 July 1991.

dynamics of the thinning process involves qualitatively different physics. For a start, the total surface area does not change; and one must also account for the influence of the surfactant molecules (layers of which at the air–water interfaces stabilize the film in the first place). At low surfactant concentrations, the canonical theory of colloidal forces, DLVO theory⁶, predicts two types of equilibrium film that differ in thickness: a common black film (10–100 nm) and a Newton black film (~5 nm). Mechanical disturbances can trigger thinning of the former to the latter.

When, however, the surfactant concentration is increased above the critical micelle point, the surfactant molecules aggregate into spherical micelles with their hydrophobic tails buried inside. The thinning process then occurs in a stepwise manner which can be observed as the nucleation and growth of dark (thinner) spots on the film surface^{7,8} (P. A. Kralchevsky, University of Sofia). The steps are explained in terms of the packing of micelles, which adopt a layered configuration in films only a few micelle diameters in thickness. Calculations (C. J. Radke, University of California, Berkeley) show that the micelles can be regarded as essentially hard-sphere-like, a proposal supported by the fact that the same effect is observed in films containing a suspension of much larger latex spheres (Kralchevsky).

This stepwise thinning recalls the oscillatory force curves measured when a liquid such as water is compressed between two mica plates⁶ (J. Israelachvili, University of California, Santa Barbara) — films only a few molecular diameters thick become stratified, and jumps in the inter-plate distance occur as successive layers are squeezed out. But the latter are equilibrium, first-order transitions, whereas the thinning of micellar soap films takes place spontaneously at constant pressure, each intermediate thickness being only metastable. The latter process is, in other words, a dynamical one.

The surface-force apparatus⁹ described by Israelachvili can be said with

justification to have revolutionized the measurement of short-range intermolecular forces. It is also now finding applications for dynamical studies. J. Klein (Weizmann Institute) reported on the effect of shear on the forces between two mica sheets coated with polymers (either grafted or adsorbed)¹⁰. By suspending the sheets on leaf springs, oscillatory tangential motion is possible, imposing a shear on the fluid between the plates. The lateral ('rubbing') force between the plates is affected remarkably little by shear, even at separations rather less than twice the effective length

of the polymer chains. But the repulsive force normal to the plates increases when the shear velocity exceeds a critical value, whereas one might expect a decrease as the dangling chains get combed down. The polymer coats thus help to keep the plates apart under shear. It seems that adsorption of polymers has been given little consideration as a means of lubrication; but perhaps here the physics is for once telling chemical engineers something that empiricism previously has not. □

Philip Ball is an assistant editor of Nature.

POPULATION ECOLOGY

Chaos in time and space

Anthony R. Ives

ECOLOGISTS' hopes of explaining the population dynamics of plants and animals in terms of simple, stable interactions among species were shattered by the discovery in the mid 1970s that even simple interactions may result in chaotic population dynamics¹. On page 255 of this issue², M. P. Hassell, H. Comins and R. M. May put to rest any such hopes that linger on — they show that simple interactions among species may lead not only to chaotic fluctuations through time, but also chaotic patterns in space where population booms and busts occur in a seemingly random pattern.

Morphogenesis

The idea that simple processes may lead to complex spatial patterns entered the field of biology in 1952. In his classic paper on morphogenesis, Alan Turing³ demonstrated that simple chemical events can produce stable spatial patterns in chemical reactants. The minimum requirement is two chemicals, one an activator that stimulates the production of the second, which itself inhibits the first. Then, with certain diffusion rates, stationary wave-like patterns evolve in which ridges in the inhibitor concentration coincide with troughs in the activator concentration. This mathematical result has profound consequences for biological morphogenesis because it provides an explanation for pattern formation during the development of an initially homogeneous collection of cells, such as is found in an embryo.

Turing's model is closely akin to the ecological model considered by Hassell *et al.*, which addresses the interactions between a parasitoid and its host. (A parasitoid is a predator, typically an insect, which lays offspring in its host; the offspring then kill and consume the host.) Here, the activator is the host,

and the inhibitor is the parasitoid. Both host and parasitoid diffuse throughout the environment which, for the sake of the mathematical calculations, is divided up into identical, contiguous patches. Thus this ecological model has all the ingredients needed to produce spatial pattern. The surprise that the spatial patterns are chaotic, however, comes from the type of interaction between parasitoids and hosts.

In Turing's model, the reaction between chemicals is stable; if the chemicals are confined to a small area so that diffusion is not an issue, the chemicals will reach stable concentrations. In Hassell *et al.*'s ecological model, the interaction between parasitoids and hosts is unstable; confining parasitoids and hosts to a small area will lead to the extinction of one or both. This is a natural assumption, because many experiments have shown instability in predator–prey interactions. With instability between the activator and the inhibitor, Turing's stable spatial patterns are replaced by a plethora of complex patterns, including spiral waves and true spatial chaos. Such exotic patterns seem to be the rule, with better behaved spatial patterns being the exception. Patterns like these are new to ecological models, because previous models assume (unrealistically, for most species) that hosts and parasitoids disperse evenly throughout a large region, rather than diffuse only among adjacent areas.

What is the evidence that chaotic spatial patterns exist in natural populations? The first place to look is controlled laboratory experiments that guarantee an initially homogeneous environment. For example, the patterns produced during Carl Huffaker's studies on the interactions between predatory and prey mites⁴ look suspiciously like chaos. In his experiments, decreasing the freedom with which predatory mites could dis-

1. Ondarçuhu, T. & Veyssié, M. *Nature* **352**, 418–420 (1991).
2. Redon, C., Brochard-Wyart, F. & Rondelez, F. *Phys. Rev. Lett.* **66**, 715–718 (1991).
3. Huppert, H. E. *Nature* **300**, 427–429 (1982).
4. Melo, F., Joanny, J. F. & Fauve, S. *Phys. Rev. Lett.* **63**, 1958–1961 (1989).
5. Cazabat, A. M., Heslot, F., Troian, S. M. & Carles, P. *Nature* **346**, 824–826 (1990).
6. Israelachvili, J. N. *Intermolecular and Surface Forces* (Academic, London, 1985).
7. Nikolov, A. D. & Wasan, D. T. *J. Colloid Interface Sci.* **133**, 1–12 (1989).
8. Nikolov, A. D., Kralchevsky, P. A., Ivanov, I. B. & Wasan, D. T. *J. Colloid Interface Sci.* **133**, 13–22 (1989).
9. Israelachvili, J. N. & Adams, G. *J. chem. Soc. Faraday Trans. 1* **74**, 974–1001 (1978).
10. Klein, J., Perahia, D. & Warburg, S. *Nature* **352**, 143–145 (1991).
11. Widom, B. *J. chem. Phys.* **84**, 6943 (1986).

perse among patches in an experimental arena, and increasing the number of patches, allowed predators and prey to coexist for over 200 days, whereas unobstructed predators would wipe out the prey in less than 32 days. Although considerable theoretical work has been devoted to explaining how Huffaker's mites persisted, no one has looked carefully at the resulting spatial patterns. If these patterns are in fact chaotic, this provides support for Hassell *et al.*'s striking conclusion that even chaotic patterns lead to persistence when there are enough patches.

Natural settings

Detecting spatial chaos in natural settings is much harder. Several host-parasitoid systems have many of the requisite features for spatial chaos, and complex spatial patterns in population densities have been observed. For example, the larch sawfly appears to become extinct locally after experiencing high parasitism rates by the wasp *Olesicampe benefactor*, but then reinvades after the wasp population crashes⁵. This produces a patchwork of presence and absence of larch sawflies as they play an ecological shell game with their predator. But considerable difficulties arise in distinguishing patterns resulting from the interaction among hosts and parasitoids, and those resulting from external events, such as sudden weather changes. Disentangling them is a challenge that must be faced by experimentalists and theoreticians together, because it will almost certainly require fresh approaches to analysing data.

A further complication stems from the curious relationship between spatial patterns generated through the interaction between hosts and parasitoids, and spatial heterogeneity in the environment. Models with a spatially heterogeneous environment⁶⁻⁹ show that when there are differences in demographic parameters for either host or parasitoid in different patches, the dynamics of the combined patches are usually more stable than those of any patch alone. This results because each patch cannot maintain its characteristic dynamics when tied to the asynchronous dynamics of the others. If this observation also holds for systems such as those considered by Hassell *et al.*, then spatial heterogeneity in the environment might actually serve to counteract the spatial heterogeneity

associated with spatial chaos.

The fundamental implication of the work by Hassell *et al.* is that explaining the temporal and spatial patterns of population dynamics may be very difficult, because even simple interactions lead to complicated patterns. Although this might frustrate many ecologists, it is really a cause for hope. We know that

MILKY WAY GALAXY

Annihilation near the centre

R. Ramaty and R. E. Lingenfelter

AMIDST the violence of the centre of our Galaxy there appears to be a bright spot (or perhaps several) where electrons are annihilated in collisions with their antiparticles, positrons, creating highly characteristic γ -rays. Observations from the Soviet satellite GRANAT last year and earlier this year show that the brightest source of hard X-rays in the Galactic Centre is a source of such annihilation radiation. The demonstration by Bally and Leventhal on page 234 of this issue¹, that this X-ray hotspot is associated with (perhaps embedded within) a molecular cloud containing the equivalent of 10^5 solar masses, suggests that the annihilation source could be powered by accretion of interstellar material onto a stellar-mass black hole. This may turn out to be the first example of a compact object, perhaps a black hole, feeding on gas in a molecular cloud.

Positron annihilation radiation from the direction of the Galactic Centre has been observed for over two decades. It was first detected² in 1970 and 1971 by balloon-borne instruments, but only in 1977 was the energy of the annihilation line (expected to equal the electron's rest-mass energy, 511 keV) accurately determined³. The observed line-centre energy, 510.71 ± 0.5 keV, clearly established that the radiation is due to positron annihilation.

Subsequent satellite- and balloon-borne instruments showed that the radiation's intensity could change dramatically over periods of less than half a year. The line flux decreased after 1979, increased between 1984 and 1988, and possibly decreased again in 1989. Two recent balloon-borne observations showed⁴ changes between May and October 1988. The rapidity of these changes implies that a significant fraction of the observed 511-keV flux comes from one or perhaps a few compact sources in the central region of our Galaxy. We have shown⁵ that the positrons in such compact sources are most probably produced in interactions of γ -rays with γ -rays, and, if this production occurs near the inner regions of the accretion disks of black holes, then their

the spatial patterns of many organisms are complex. Hassell *et al.* tell us that the processes underlying these patterns may nonetheless be simple, and therefore we should look for them. □

Anthony R. Ives is in the Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706, USA.

masses must be less than several hundred solar masses.

The observed 511-keV flux also depends on the opening angle — or field of view — of the detectors. This first became evident from the comparison of two observations in 1977 with balloon-borne detectors, one with a 100° field of view, the other 15° . Although carried out within 10 days of one another, the observations from the detector with the larger field of view showed a significantly larger flux than those from the 15° instrument. More strikingly, the 130° detector on the Solar Maximum Mission satellite revealed⁶ large fluxes of 511-keV emission in 1980–88, while balloon-borne detectors of similar sensitivity but with a field of view of only 15° yielded only upper limits in 1981 and 1984. The disparity suggests that at least a fraction of the observed annihilation radiation from the Galactic Centre is spatially distributed. We have proposed⁷ a two-component model according to which the total 511-keV emission is a superposition of a steady, distributed component (the result of positrons from the radioactive decay of the products of galactic nucleosynthesis) and a variable one from compact sources.

As mentioned above, the positrons in the compact object (or objects) could be produced as electron-positron pairs by γ -rays interacting with each other. Continuum observations at photon energies greater than 511 keV show that for the variable compact object near the Galactic Centre, the energy available for pair production at the source could be about 30 MeV per pair, a value well in excess of the minimum of 1.02 MeV required if the pairs are produced by photons colliding head-on. To achieve the desired rate of pair production, the density of the photons must be sufficiently high, and this constrains the size of the source to less than 10^4 km. Such a region of high photon density could exist in the vicinity of a black hole of mass less than several hundred solar masses.

This black hole may in fact be the X-ray source 1E1740.7-2942 about 0.9° from the Galactic Centre, as annihilation

1. May, R. M. *Nature* **261**, 459–467 (1976).
2. Hassell, M. P., Comins, H. & May, R. M. *Nature* **353**, 255–258 (1991).
3. Turing, A. M. *Phil. Trans. R. Soc.* **B237**, 37–72 (1952).
4. Huffaker, C. B. *Hilgardia* **27**, 343–383 (1958).
5. Murdoch, W. W., Chesson, J. & Chesson, P. L. *Am. Nat.* **125**, 344–366 (1985).
6. Reeve, J. D. *Am. Nat.* **132**, 810–836 (1988).
7. Taylor, A. D. *Ecology* **71**, 429–433 (1990).
8. Murdoch, W. W., Briggs, C. J., Nisbet, R. M., Gurney, W. S. C. & Stewart-Oaten, A. *Am. Nat.* (in the press).
9. Ives, A. R. *Am. Nat.* (in the press).

radiation from this source was apparently observed^{8,9} with the SIGMA detector on the Soviet satellite GRANAT. SIGMA is an imaging X- and γ -ray spectrometer capable of providing images in the energy range 35 keV to 1.3 MeV with an angular resolution of approximately 15 arcminutes over a field of view of about 16 square degrees. Observations made this year and last year show that in this energy range the sky around the Galactic Centre is dominated by two sources, 1E1740.7-2942 and GRS1758-258. The first source, at a distance of at least 120 parsec from the dynamic hub of the Galaxy, has been known for some time as the strongest hard X-ray source in the Galactic Centre region^{10,11}. GRS1758-258 is a newly discovered source, which in past observations was confused with another X-ray source, GX5-1.

During most of the spring and autumn of 1990, the hard X-ray spectrum of 1E1740.7-2942 was in good agreement with that expected¹² from an accretion disk model around a black hole with the X-rays energized by Compton scattering. Such a spectrum is characterized by a negative slope at energies above several tens of keV. But data on 13–14 October 1990 revealed a soft γ -ray feature with a positive slope (its flux increased with increasing photon energy) from 200 keV to a cutoff at about 700 keV. Such a feature suggests the presence of a relativistic plasma¹³ consisting of electrons and positrons. No such feature has been observed in the spectrum of GRS1758-258. It is thus tempting to associate 1E1740.7-2942 with the variable source of 511-keV emission.

This association, however, is still uncertain because the variable, narrow 511-keV line, which was previously seen⁴ with the germanium detectors, is not obvious in the spectrum of 1E1740.7-2942. The poor energy resolution of the SIGMA detectors is the main cause of this uncertainty. Indeed, while the positive-slope spectrum of 1E1740.7-2942 is well fitted by a very broad line component due to annihilation in redshifted hot gas near the black hole, the presence of a narrow 511-keV line cannot be ruled out¹⁴.

A narrow line would be produced by positrons streaming away from the black hole and annihilating in colder gas (hence the small width) far from the hole. If the positive-slope spectrum of 1E1740.7-2942 is indeed due to a narrow line, the rapid variability implied by the 2-day observations can be used to infer the density of the gas, which must be greater than 10^7 cm^{-3} . Although this is 100 times more dense than the average expected in the cloud observed by Bally and Leventhal¹, these authors suggest that near the compact object the accreting gas may reach the required density.

But it is not clear that the velocity of such gas would be consistent with the observed line-centre energy and line width.

It is of course also possible that the annihilating positrons are produced in 1E1740.7-2942 on timescales of the order of a day, but that the narrow line is formed on longer timescales at a site removed from the hole. Such a site may in fact be the molecular cloud observed by Bally and Leventhal. In this picture, the 1-day activity of 1E1740.7-2942 could produce a large number of electron-positron pairs, a fraction of which could escape into the cloud. With an ambient gas density of 10^5 cm^{-3} , the positrons could annihilate and form a narrow 511-keV line on a timescale of about 1 year.

In addition to 1E1740.7-2942, recent observations have revealed variable annihilation features also from nova Muscae¹⁵ and an object along a line of sight near the cataclysmic variable V1223 Sagittarii¹⁶. Although these objects are outside the fields of view of the balloon-borne detectors, their discovery suggests that there may be a whole new class of compact sources of annihilation radiation which produce variable narrow 511-keV line emission at various times. The origin of the observed line emission thus remains unresolved. What is needed is a systematic mapping of the Milky Way with detectors of high-spatial and high-spectral resolution, which could localize the sources and accurately determine their energy spectra. $\square$

R. Ramaty is in the Laboratory for High Energy Astrophysics, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

R. E. Lingenfelter is in the Center for Astrophysics and Space Sciences, University of California San Diego, La Jolla, California 92093, USA.

1. Bally, J. & Leventhal, M. *Nature* **353**, 234–237 (1991).
2. Haymes, R. C. et al. *Astrophys. J.* **201**, 593–602 (1975).
3. Leventhal, M., MacCallum, C. J. & Stang, P. D. *Astrophys. J.* **225**, L11–L14 (1978).
4. Gehrels, N. et al. *Astrophys. J.* **375**, L13–L16 (1991).
5. Lingenfelter, R. E. & Ramaty, R. in *The Galactic Center* (eds Riegler, G. R. & Blandford, R. D.) 148–159 (Am. Inst. Phys., New York, 1982).
6. Share, G. H. et al. *Astrophys. J.* **326**, 717–732 (1988).
7. Lingenfelter, R. E. & Ramaty, R. *Astrophys. J.* **343**, 686–695 (1989).
8. Paul, J. et al. in *Gamma Ray Line Astrophysics* (eds Durouchoux, Ph. & Prantzos, N.) 17–28 (Am. Inst. Phys., New York, 1991).
9. Sunyaev, R. et al. in *Gamma Ray Line Astrophysics* (eds Durouchoux, Ph. & Prantzos, N.) 29–41 (Am. Inst. Phys., New York, 1991).
10. Skinner, G. K. et al. *Nature* **330**, 544–547 (1987).
11. Cook, W. R. et al. in *The Center of the Galaxy* (ed. Morris, M.) 581–585 (Kluwer, Dordrecht, 1989).
12. Sunyaev, R. & Titarchuk, L. *Astr. Astrophys.* **86**, 121–138 (1980).
13. Liang, E. P. & Dermer, C. D. *Astrophys. J.* **325**, L39–L42 (1988).
14. Bouchet, L. et al. *22nd Int. Cosmic Ray Conf. Papers* (in the press).
15. Sunyaev, R., Jourdain, E. & Goldwurm, A. *IAU Circ. No.* 5201 (1991).
16. Briggs, M., Matteson, J., Gruber, D. & Peterson, L. *IAU Circ. No.* 5229 (1991).

A sideways move

AMONG the many consequences that Henry Ford never imagined when he began to mass-produce cars, is the parking problem. It now dominates our cities. Their once spacious streets are reduced to choked crawlways, and their noble centres are disfigured with those most unlovely creations of modern architecture, elevated car-parks. Worse still, most parking space is wasted. A filled parking building contains less than 10 per cent of its internal volume as cars; a parking lot or roadside is seldom packed to more than 50 per cent density. The rest of the space is needed for access. So Daedalus began dreaming up ways of close-packing parked cars: systems of racks or shelving in which they could be placed and retrieved by fork-lift trucks, or spiralled carousel-type storage conveyors whose rotation could return any given car to the exit point on command. But he now has a simpler solution. He is devising a car which can go sideways.

The plan is very neat. The tyres of some big trucks have rotating air-connections to the chassis, enabling the driver to monitor their pressures continuously, and even inflate them from a pump on the vehicle. Daedalus is adapting this system to work his special 'Sideways Tyre', which has three inner tubes side by side. In normal driving, all three are maintained at the proper pressure. But with the vehicle stationary, they can be individually deflated by a special valve system. If the outer tube goes flat, the wheel leans a little in that direction. The collapse of the middle one increases that lean. If the outer tube is then reflat to grip the road again while the inmost tube is deflated, the tyre is held in its shifted position. Re-inflation of the other two tubes stabilizes it there. By rapidly repeating this cycle, the tyre shuffles peristaltically sideways caterpillar-fashion.

Sideways Tyres and their associated pneumatic plumbing can be fitted fairly simply even to existing cars, transforming their mobility. If the front tyres go left while the back ones go right, the car rotates on its own axis; a wonderful way of getting into, out of, and round tight corners. If all tyres creep the same way, the car goes sideways. Parking in even the narrowest space is easy: just draw up alongside, and slide in sideways.

In the vicious parking battle of our choked cities, Sideways Tyres will confer a massive selective advantage. A car fitted with them can block in other cars, but cannot be blocked in itself. So once introduced, Sideways Tyres should spread rapidly, until every car has them. Roads and parking lots will at last be efficiently close-packed, and their capacity will be doubled.

DAVID JONES

Coelacanth's relationships

SIR — Gorr *et al.*¹ claim that analyses of α - and β -haemoglobin sequences from the coelacanth *Latimeria chalumnae* indicate that coelacanths are the closest living relatives of tetrapods. Forey² noted that this conclusion is contradicted by results from other molecules and that a combined analysis of mitochondrial DNA and haemoglobin data favours a closer relationship of lungfishes and tetrapods. We believe that even the haemoglobin data considered alone do not support a close relationship between coelacanths and tetrapods.

Gorr *et al.* observed from pairwise comparisons that haemoglobins of bony fishes are generally more similar to larval than to adult amphibian haemoglobins, which was taken as evidence for orthology³ between bony-fish and larval-amphibian (tadpole) haemoglobins. To identify the closest tetrapod relative, bony-fish haemoglobins were compared to those of tadpoles. Coelacanths were most similar to tadpoles for the β -chain, whereas teleosts were most similar to tadpoles for the α -chain. These results were corroborated by phenetic (CLUSTAL⁴) and phylogenetic (PROTPARS⁵) clustering algorithms. Gorr *et al.* attempted to resolve the conflict between the α and β results by using a "cladistic concept" that identifies "synapomorphous positions" as those for which a particular bony fish and at least one tadpole share a residue not found in any other bony fish or in humans. According to this new criterion, a tadpole-coelacanth relationship was favoured for both chains.

Their cladistic method is seriously flawed in that a simple count of 'shared derived' residues may give a misleading impression of the relative support for alternative phylogenies. For the example in Fig. 1, as long as at least one charac-

ter of each type (I and II) is present, tree *a* will always be preferred under the parsimony criterion⁶, as each character has been explained by the minimum possible number of amino-acid changes.

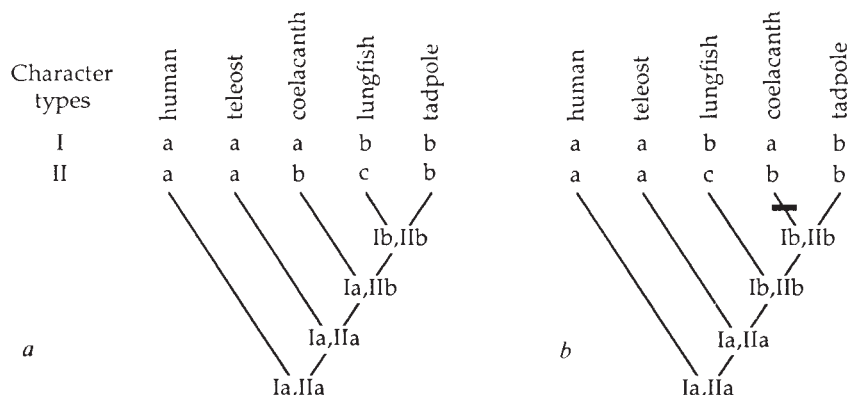


FIG. 1 Demonstration that the method of counting exact matches of (inferred) derived residues described in ref. 1 provides misleading support for certain groupings. Suppose that each character is distributed according to one of the two patterns referred to as type I or type II. Tree *a* is maximally parsimonious for the two character types considered jointly, whereas tree *b* requires either a reversal (indicated by bar) or a parallelism in characters of type I. Internal node labels represent ancestral character states inferred under the parsimony criterion.

Any other tree (for example, tree *b*) will require at least one instance of homoplasy (convergence, parallelism, or reversal) to explain the data. Yet if there is even one more character of type II than of type I, the 'cladistic' method of Gorr *et al.*¹ will prefer tree *b* (or any other tree that groups coelacanth with tadpoles) over tree *a*, at the expense of unnecessary homoplastic events. Thus, their method clearly is not 'cladistic'. A further problem is created by the unequal representation of taxa in the three bony-fish groups being compared to tadpoles. In order for a residue found in any one of the four teleost species to be counted as a synapomorphy, it must be absent from the remaining ones.

Because teleosts are (presumably) monophyletic, any position for which all represented teleosts have retained the residue found in their common ancestor would not be a 'synapomorphy' with tadpoles, although such conserved positions might ordinarily be considered more reliable. On the other hand, coela-

cans and lungfishes are represented by a single species each, and would thus be expected to have higher synapomorphy counts. When Gorr *et al.* corrected for the problem of unequal representation (column 3 of their Table 2), β -haemoglobin still supported a coelacanth-tadpole relationship, but α -haemoglobin supported a teleost-tadpole grouping, in agreement with their earlier analyses. It is puzzling that they chose to ignore this result entirely in their discussion.

Because Gorr *et al.* placed greater emphasis on the β -chain results, we investigated the strength of support for the conclusions derived from that data set. Their CLUSTAL tree, based on 20 β -globin sequences, contains the main groupings shown in Fig. 2*a*; they claimed to have obtained the same groupings with PROTPARS. Even the authors of CLUSTAL advise against interpreting its output as a phylogeny⁴; consequently, we have examined the support for their topology using parsimony methods. We obtained the globin sequences listed in their Fig. 2 from databanks and aligned them using the progressive alignment method of Feng and Doolittle⁷ and using CLUSTAL (a "quick and dirty" version of Feng and Doolittle's algorithm⁴). The alignments obtained had minor differences between each other as well as with the partial alignment presented in Fig. 1 of Gorr *et al.* (presumably obtained by CLUSTAL). The results we present here are based on the alignment obtained by Feng and Doolittle's method, although comparable analyses based on various

TREE TOPOLOGIES AND LENGTHS FROM PAUP FOR β -HAEMOGLOBIN DATA

TOPOLOGY	LENGTH	
	PROTPARS	FITCH
((lungfish,adult),(coelacanth,(teleost,larval))) ^b	951	753
(coelacanth,(larval,(teleost,(lungfish,adult)))) ^c	951	743
((coelacanth,larval),(adult,(lungfish,teleost)))	952	743
(larval,(coelacanth,(teleost,(lungfish,adult))))	953	745
(coelacanth,(larval,(adult,(lungfish,teleost))))	953	739
(larval,(coelacanth,(adult,(lungfish,teleost))))	955	742
(adult,(lungfish,(teleost,(coelacanth,larval)))) ^a	956	755
(adult,((coelacanth,lungfish),(teleost,larval)))	956	760
(larval,(coelacanth,(lungfish,(teleost,adult))))	956	747

Lengths are shown for best-fitting trees matching the relationships indicated by each topology for the full set of 20 sequences under the PROTPARS⁵ and Fitch⁶ (unordered) criteria for counting amino-acid changes, respectively. Trees were rooted by inclusion of shark sequences as an outgroup. Superscripts refer to trees in Fig. 2. We did not examine all 105 possible topologies for the major groups; other topologies of similar length may exist.

combinations of the gap features of all three alignments gave very similar results.

We conducted phylogenetic analyses using PAUP⁸ and a change-counting criterion identical to that employed by PROTPARS. Using a more extensive heuristic search procedure than the one implemented in PROTPARS, PAUP obtained the two trees shown in Fig. 2b,c. These trees have little in common, both in terms of the position of the coelacanth and in the pattern of gene duplication required to produce larval and adult amphibian genes. Neither tree supports a sister-group relationship between coelacanth and tadpole haemoglobins. The most parsimonious tree consistent with the groupings of Gorr *et al.* (Fig. 2a), obtained using the 'constraints' feature of PAUP, was five steps longer than our most parsimonious trees. Although Gorr *et al.* claim that the relationships implied by Fig. 2a were consistently obtained using various combinations of taxa, we were able to obtain this tree as most parsimonious only by omitting the shark sequences. However, removing the only clear outgroup taxa forces the use of questionable assumptions regarding the orthology of the

larval-amphibian and fish haemoglobin genes to obtain a rooted tree.

We do not suggest that either of our trees (Fig. 2b,c) represents the true phylogeny of vertebrate β -haemoglobin genes. In fact, a wide variety of trees can explain the data nearly as well (see table). For example, only two additional steps were required to obtain a topology consistent with the orthology of fish and adult-amphibian haemoglobins (rather than fish and larval-amphibian haemoglobins, as suggested by Gorr *et al.*). That trees differing radically in topology require very similar numbers of steps with the β -haemoglobin data indicates to us that there is no support for the conclusion of Gorr *et al.* about coelacanth-tetrapod relationships or even the pattern of β -haemoglobin gene duplication in vertebrates.

DAVID W. STOCK

Department of Ecology, Ethology, and Evolution,

University of Illinois,

505 S Goodwin Ave,
Urbana, Illinois 61801 USA

DAVID L. SWOFFORD

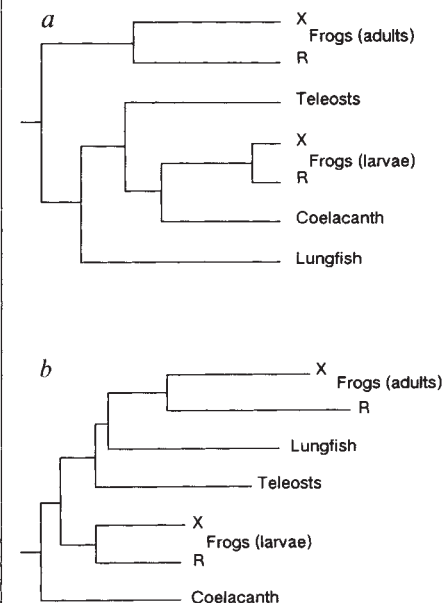
Center for Biodiversity,
Illinois Natural History Survey,
607 E Peabody Drive,
Champaign, Illinois 61820 USA

1. Gorr, T., Kleinschmidt, T. & Fricke, H. *Nature* **351**, 394–397 (1991).
2. Forey, P. *Nature* **351**, 347–348 (1991).
3. Fitch, W. M. *Syst. Zool.* **19**, 99–113 (1970).
4. Higgins, D. G. & Sharp, P. M. *Gene* **73**, 237–244 (1988).
5. Felsenstein, J. *PHYLIP Manual Version 3.2* (University Herbarium, University of California, Berkeley, 1989).
6. Fitch, W. M. *Syst. Zool.* **20**, 406–416 (1971).
7. Feng, D.-F. & Doolittle, R. F. *Meth. Enzym.* **183**, 375–387 (1990).
8. Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony* (Illinois Natural History Survey, Champaign, Illinois, 1991).

SIR — Which among coelacanths, lungfish or teleosts are our closest relatives? Puzzles involving evolutionary relationships are now most likely to be solved using DNA or protein sequence data, which can be reliable if used in sufficient quantity and analysed by the most appropriate techniques. In their analysis of haemoglobin sequences, Gorr *et al.*¹ conclude that the coelacanth *Latimeria* is the most closely related to tetrapods. Their analyses are rendered suspect by unequal rates of evolution; more appropriate techniques yield results that do not contradict the suggestion from other molecular data that lungfish are the closest living relatives of tetrapods².

The tree presented by Gorr *et al.*¹ (see figure) was derived from analysis of β -globin sequences using the CLUSTAL program, in which a dendrogram is produced as a guide for multiple alignment. We have said³ that these dendrograms "are not intended to be used as phylogenetic trees" because (1) the sequence comparison scores used are "only crude estimates"; and (2) the method of phy-

logenetic inference used, UPGMA, assumes equal rates of evolution — this assumption, when violated, "may lead to errors in the tree topology". There are clear indications that rates of β -globin



Phylogenetic trees for β -globin sequences. Among frogs X is *Xenopus* and R is *Rana*. The trees were rooted using shark outgroups. Eighteen sequences were used in all, but some branches within obvious monophyletic groups were pruned for clarity. a, UPGMA analysis of protein similarity scores by CLUSTAL³; b, neighbour-joining analysis⁵ of corrected amino-acid divergence values⁴.

evolution have been different in different lineages, and we believe the conclusions reached by Gorr *et al.* are thus an artefact.

The rate differences can be seen by taking the number of amino-acid differences between sharks and the other groups. Using either shark as the outgroup, the coelacanth and larval frog (tadpole) β -globins exhibit fewer differences from the shark than do lungfish, teleost and adult frog globins (see table). These values understate the disparity in rates, because they have not been corrected for multiple hits, and each includes a long common branch to the shark. For example, the least-squares estimates of

β -GLOBIN AMINO-ACID SEQUENCE DIFFERENCES		
	Versus shark*	
	(1)	(2)
Frog† (adult)	92	83
Frog† (larval)	78	69
Coelacanth	76	72
Lungfish	90	80
Teleost‡	87	87

* (1) Port Jackson shark, (2) dogfish (*Squalus*).

† *Xenopus laevis*, taken as typical.

‡ *Salmo irideus*, taken as typical.

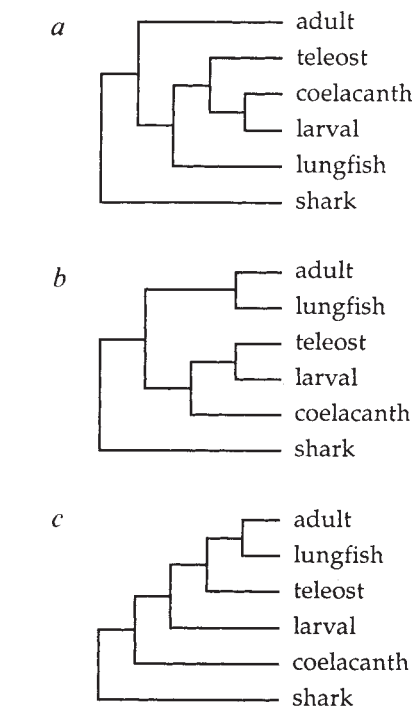


FIG. 2 Topologies obtained for β -haemoglobin by Gorr *et al.*¹ using CLUSTAL and PROTPARS (a) compared with most parsimonious trees obtained in our analyses using PAUP under the PROTPARS criterion (b,c). Although we used 20 sequences in our analyses, only the major groupings are shown. The categories 'adult' and 'larval' refer to amphibian sequences. Heuristic searches were performed using 10 random addition sequences and TBR branch swapping⁸.

the branch lengths to coelacanth and lungfish from their common ancestor are 0.27 and 0.65 amino-acid replacements per site, respectively, suggesting that the average rate of β -globin evolution in the lungfish lineage has been more than twice that in the coelacanth lineage, since their divergence. (These particular values come from comparisons with the Port Jackson shark as outgroup, with correction for multiple replacements⁴.) Comparisons between *Xenopus* and *Rana* of the β -globins expressed in larvae (average 44.7 amino-acid differences) and in adults (average 62.4) emphasize that the tadpole β -globins have also evolved slowly.

The neighbour-joining method⁵ is more appropriate than UPGMA, because it allows for unequal rates of evolution, and appears to be one of the most efficient algorithms for estimating the correct phylogeny⁶. This method suggests that the adult frog β -globins are orthologous to the fish proteins, and clusters lungfish with adult frogs (see figure), which is consistent with a result obtained from mitochondrial DNA². But this topology is not statistically reliable, as lungfish and frogs cluster in less than 60% of 1,000 bootstrap resamples. Thus, there are insufficient data to make firm phylogenetic statements. Gorr *et al.* conclude that in frogs the larval globin rather than adult form is orthologous to the fish sequences. Because the tadpole and coelacanth branches are both short, their observation that the β -globin of *Latimeria* is the most similar to tadpoles arises because both have undergone slow recent evolution, and not because they are the most closely related. Gorr *et al.* also used the maximum parsimony algorithm to support their conclusions, but this method also has a tendency to artefactually cluster long branches with long branches (and thus short with short)⁷.

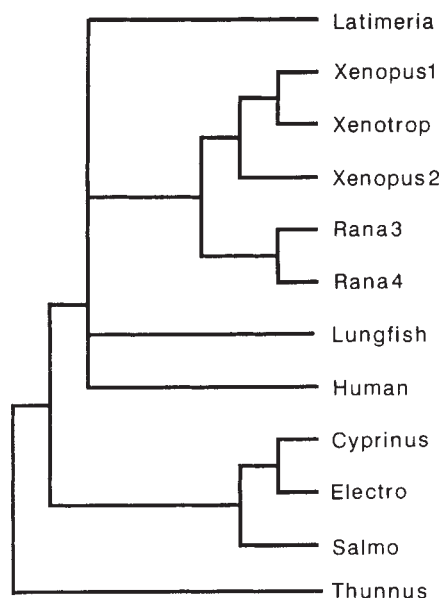
PAUL M. SHARP
ANDREW T. LLOYD

Department of Genetics,
Trinity College,
Dublin 2, Ireland

DESMOND G. HIGGINS
European Molecular Biology Laboratory,
6900 Heidelberg, Germany

1. Gorr, T., Kleinschmidt, T. & Fricke, H. *Nature* **351**, 394–397 (1991).
2. Meyer, A. & Wilson, A. C. *J. molec. evol.* **31**, 359–364 (1990).
3. Higgins, D. G. & Sharp, P. M. *Gene* **73**, 237–244 (1988).
4. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge University Press, 1983).
5. Saitou, N. & Nei, M. *Molec. biol. Evol.* **4**, 406–425 (1987).
6. Saitou, N. & Imanishi, T. *Molec. biol. Evol.* **6**, 514–525 (1989).
7. Felsenstein, J. *Syst. Zool.* **27**, 401–410 (1978).

SIR — Gorr *et al.*¹ claim to have found that the coelacanth is more closely related to tadpoles than are lungfishes. Their Fig. 2 contains a number of surprising and unprecedented groupings.



Evolutionary tree (PAUP, strict consensus⁷) based on analyses of amino acids and DNA data of haemoglobin chains. Names can be inferred from ref. 1.

Only one previous study², which was not based on cladistic approaches, aligned teleost fishes closely with the tetrapods. The new results¹ are at odds with morphological³ and molecular^{4,5} studies.

Several obvious problems in the analyses of their data prompted us and others (see the letter from Stock and Swofford, above) to re-analyse the data of Gorr *et al.* using parsimony⁷. To avoid potential problems of differences in alignment, only data presented in Fig. 1 of Gorr *et al.* have been used in our reanalysis. Ray-finned fishes are used as outgroup in agreement with virtually all phylogenetic hypotheses, which recognize that these fishes are only remote relatives of sarcopterygians³.

In our α -chain analysis we found three equally short evolutionary trees (TL=401, CI=0.803); the strict consensus of these three equally likely solutions is shown in the figure. The α -chain does not allow one to determine the relationships among sarcopterygian groups considered (human, tadpoles, coelacanth and lungfish). In our β -chain analysis we found three equally short trees (TL=427, CI=0.787), the consensus of which groups *Latimeria* closer to the larval, frog's haemoglobin. However, a trifurcation exists with that group and the human and the lungfish. The results of our reanalysis do not agree with the reported outcomes of Gorr *et al.*, even when the sequence data from sharks⁶ were included as outgroups. Reanalyses of amino acids demonstrate that neither the α - nor the β -chains of haemoglobin permit the inferences made in ref. 1.

DNA sequences contain more information than protein sequences, because the replacement of an amino acid

by another at a certain site may sometimes require more than one nonsynonymous substitution at the DNA level. Protein sequences of the β -chains were reverse-translated into DNA (IUPAC-IUB code) and subjected to parsimony analyses⁷. Two equally short trees (TL=514, CI=625) were found, the consensus of which is shown in the figure. Here, also, there is no support for the claims of Gorr *et al.*¹.

AXEL MEYER

Department of Ecology and Evolution,
State University of New York,
Stony Brook, New York 11794, USA

ALLAN C. WILSON*

Division of Biochemistry and
Molecular Biology,
University of California,
Berkeley, California 94 720, USA

*Deceased.

1. Gorr, T. *et al.* *Nature* **351**, 394–397 (1991).
2. von Wahlert, G. *Latimeria und die Geschichte der Wirbeltiere* (Fischer, Stuttgart, 1968).
3. Forey, P. *Nature* **336**, 727–732 (1988).
4. Meyer, A. & Wilson, A. C. *J. molec. Evol.* **31**, 359 (1990).
5. Forey, P. *Nature* **351**, 347–348 (1991).
6. Kleinschmidt, T. & Sgourous, J. G. *Biol. Chem. Hoppe-Seyler* **368**, 579–615 (1987).
7. Swofford, D. L. *Phylogenetic Analysis Using Parsimony*, Version 3.0 Computer program and documentation, Illinois Natural History Survey, Urbana.

No limit to global warming?

SIR — Ramanathan and Collins¹ have proposed a mechanism which limits the warming of the tropical Pacific Ocean in the present climate. I would like to make two points. First, atmospheric water vapour and cloud, and hence greenhouse trapping and cloud forcing, tend to be reduced in regions of atmospheric subsidence and enhanced in regions of ascent. This effect can be seen in Fig. 2a of ref. 1, where the rate of increase of the greenhouse effect with temperature is smaller than elsewhere between 290 and 298 K, corresponding to the regions of subsidence in the subtropics, and increases steeply as one moves to higher temperatures corresponding to the regions of ascent in the inter-tropical convergence zone. The central equatorial Pacific was a region of subsidence in April 1985 and a region of ascent in April 1987. Hence derivatives from these two years include these dynamical effects implicitly, as Ramanathan and Collins acknowledge, and they tend to enhance the more direct effects of increases in temperature. Indeed, the main region of ascent (and low-level convergence, precipitation and cloud) in the tropical Pacific appears to be linked to the region of maximum sea temperatures, rather than the magnitude of sea surface temperature *per se*². However it is not obvious that, once a region has changed from local descent to ascent, the dynamical

effects will continue to increase at the same rate for subsequent global increases in temperature such as that expected due to increases in 'greenhouse' gases.

Second, as Ramanathan and Collins imply, their mechanism does not provide an absolute limit for climatic warming. For example, increases in solar constant or CO₂ would increase their apparent upper limit to sea surface temperatures. More important, in simulations of global warming due to increases in atmospheric CO₂, moist static energy increases through the depth of the tropical troposphere (see refs 3,4) raising the apparent threshold for deep convection and so also the maximum for sea surface temperature. This, combined with the supposition

that the approach of Ramanathan and Collins may exaggerate the rate of increase of radiative effects with global temperature because it includes a contribution from local dynamics, suggests that although increases in cirrus cloud may produce a negative feedback with increases in tropical sea surface temperature, it is not as strong as Ramanathan and Collins imply.

JOHN F. B. MITCHELL

Hadley Centre for Climate Prediction
and Research,
Meteorological Office,
Bracknell RG12 2SZ, UK

1. Ramanathan, V. & Collins, W. *Nature* **351**, 27–32 (1991).
2. Barnett, T. P. et al. *J. Climate* **4**, 487–528 (1991).
3. Wetherald, R. T. & Manabe, S. *J. Atmos. Sci.* **45**, 1397–1415 (1988).
4. Wilson, C. A. & Mitchell, J. F. B. *J. Geophys. Res.* **92**, 13315–13343 (1987).

Zinc and GABA in developing brain

SIR — In the adult hippocampus, GABA, acting on GABA-A receptors, provides most of the inhibitory drive; a selective block of GABA-A receptors causes seizures. In contrast, during the first week of postnatal life, GABA-A receptors provide most of the excitatory drive, and GABA-A blockers induce an electrical silence¹. One of the striking correlates of these changes is the presence in hippocampal neurons, during the first week of postnatal life, of synchronized network-driven GABA-A mediated giant depolarizing potentials (GDP) which we believe¹ are generated by recurrent and feed-forward excitation of pyramidal cells and interneurons by GABA. Xie and Smart² suggest that GDP may be generated by endogenous zinc which, by means of a block of GABA-A receptors, increases GABA release. Although attractive, this hypothesis cannot easily be reconciled with the following observations.

Zinc has several complex effects on receptor-gated and voltage-dependent channels which could mediate its effects on GDP. It blocks the A current³ and reduces the calcium-dependent potassium current as well as the slowly inactivating voltage-dependent calcium current⁴. It increases transmitter release in the neuromuscular function, an effect probably mediated by a rise in intracellular calcium⁵, and increases AMPA-mediated currents⁶. Furthermore, in adult hippocampal slices, GABA-B receptor antagonists do not generate GDPs; a depolarizing GABA-mediated synaptic current can be generated by 4-AP which blocks the voltage dependent potassium A and D currents⁷.

The zinc-enriched mossy fibre terminals constitute the only possible source for the high concentrations of zinc required to generate GDPs (200–300 µM). The mossy fibres, however, had a delayed

development: by day 10 of postnatal life mossy fibres are developed and there are no GDPs, whereas at day 1 GDPs are present and there are essentially no mossy fibre terminals.

The parallelism suggested by Xie and Smart between the effects of zinc in rat neonatal slices and the fifth-day fits in newborn animals is hazardous because of the considerable differences in maturation stages between the two species (the first seven days of life in the rat correspond to the fetal period in man).

Xie and Smart suggest that the functional significance of GDPs may be a rudimentary form of dendritic inhibition. We prefer the hypothesis that GDPs and the depolarizing effects of GABA provide an excitatory tone to developing hippocampal neurons, reflecting a possible trophic role of GABA in growth and differentiation. There are several lines of evidence to support this suggestion: first, in neonatal neurons activation of GABA-A receptors induces a large increase in intracellular calcium; and second, block of GABA-A receptors by bicuculline retards hippocampal cell growth and neurite formation in culture⁸.

Y. BEN-ARI

E. CHERUBINI

Inserm U29,
Hôpital de Port-Royal,
123 Boulevard de Port-Royal,
75014 Paris, France

1. Ben-Ari, Y., Cherubini, E., Corradetti, R. & Gaiarsa, J. L. *J. Physiol., Lond.* **416**, 303–325 (1989).
2. Xie, X. & Smart, T. G. *Nature* **349**, 521–524 (1991).
3. Constanti, A. & Smart, T. G. *J. Physiol., Lond.* **390**, 167P (1987).
4. Sim, J. A. & Cherubini, E. *Neuroscience* **36**, 623–629 (1990).
5. Nishimura, M. *Br. J. Pharmac.* **93**, 430–436 (1988).
6. Mayer, M. L., Vicklicky, J. & Westbrook, G. L. *J. Physiol., Lond.* **415**, 329–350 (1989).
7. Perreault, P. & Avoli, M. *J. Neurophysiol.* **61**, 953–970 (1989).
8. Bargin, G., Jahchan, T., Pollard, H. & Ben-Ari, Y. *Soc. Neurosci. Abst.* (in the press).

Firefly synchrony

SIR — Ian Stewart in News and Views¹ claims that Mirollo and Strogatz² have developed a theory that can explain the synchronization of huge congregations of South East Asian fireflies (*Pteroptix malacca*) as well as hippocampal neurons, menstrual cycles and the insulin-secreting cells of the pancreas. Although Mirollo and Strogatz do prove that synchronization of oscillators occurs, their results apply to a special class of models in specific geometries, namely 'all to all' coupled-phase advance oscillators. They point out in their paper that the model is inappropriate for *P. malacca* and would better suit the responses of the non-congregating North American species *Photinus pyralis*.

Mirollo and Strogatz suggest that an oscillator will advance its phase instantaneously to a pulsatile stimulus. This immediately implies that such an oscillator cannot be entrained to lower-frequency periodic stimuli. *P. malacca* is able to entrain to periodic light pulses above and below its intrinsic frequency³ and, once it is entrained, the phase lag is almost zero: The animals actually alter their intrinsic frequency to match that of the stimulus and, once the stimulus is removed, slowly return to their original frequency. The entrainment requires many cycles of the stimulus which is not so for the pulsatile phase-advance synchrony reported in ref. 2.

The geometry of connections in ref. 2 is also somewhat unrealistic since it assumes that all oscillators are connected to all others. In congregating groups of fireflies the insects tend to space themselves out so that they are no closer than about a third of a metre apart. They will ignore the flashes of any animal that is farther than about a metre⁴. Thus any given insect sees about 100 others. In a tree with thousands of animals, the coupling is far from 'all to all'.

The real question of interest is how the animals modify their intrinsic frequencies in such a way as to achieve almost perfect synchrony. Mirollo and Strogatz make a valuable contribution in analysing collections of pulse-coupled oscillators, but Stewart's claim that "all these systems attain synchrony by the same mathematical mechanism" ignores a large body of facts about these diverse and fascinating biological phenomena.

G. BARD ERMENTROUT

Department of Mathematics,
University of Pittsburgh,
Pittsburgh,
Pennsylvania 15260, USA

1. Stewart, I. *Nature* **350**, 557 (1991).
2. Mirollo, R. & Strogatz, S. *SIAM J. appl. Math.* **50**, 1645–1662 (1990).
3. Hanson, F. E. *Fdn Proc.* **37**, 2158–2164 (1978).
4. Buck, J. & Buck, E. *Scient. Am.* **234**, 74–85 (1976).

Selling neuroscience

Susan A. Greenfield

The Enchanted Loom: Chapters in the History of Neuroscience. Edited by Pietro Corsi. Oxford University Press: 1991. Pp. 383. £50, \$60.

WE are currently celebrating 'the decade of the brain': hence *The Enchanted Loom*, the giant expansion of a catalogue to an exhibition on the development of neuroscience. The spirit of the art gallery lingers, for illustrations comprise three-quarters of the content and are probably among the most exhaustive, diverse and aesthetic ever collected together on the subject. In fact, the truly lay reader would fare best by concentrating exclusively on these arresting visual images and the accompanying self-explanatory legends.

The first section gives an account of attitudes to, and uses of, memory through the ages. To those interested in the earliest thoughts on issues that were eventually to become the basis of neuroscience, this rather idiosyncratic choice of an introductory topic will come as a disappointment. A survey of the ideas of Plato, Hippocrates and Democritus on the nature of mind would have served as a far more appropriate forerunner to the next section, which, although mentioning the Greeks in a few lines, is mainly a valuable account of theories of brain function from the time of Descartes. The narrative here is unsensationalist and strictly chronological. The only problem is that the task thus falls to the

reader to tease out the progression of parallel themes as they preoccupied, side-tracked or inspired our predecessors: the location of the soul, animistic versus mechanistic operations, the localization of function compared to more flexible holistic systems. The third section is concerned with the study of the brain as we know it today. We see how Paul Broca, Carl Wernicke, Wilder Graves Penfield and Charles Sherrington were identifying regions linked in some way to specific 'functions'. The exciting aspect of these studies is not necessarily the localization of a particular function

per se, but rather the way in which individual discoveries were used to generate principles as to how the brain worked as a cohesive whole.

Hereafter, the text ceases to be historical and spatters the reader with grapes of brief chapters that introduce topics ranging from an account of the study of the neuron through to PET

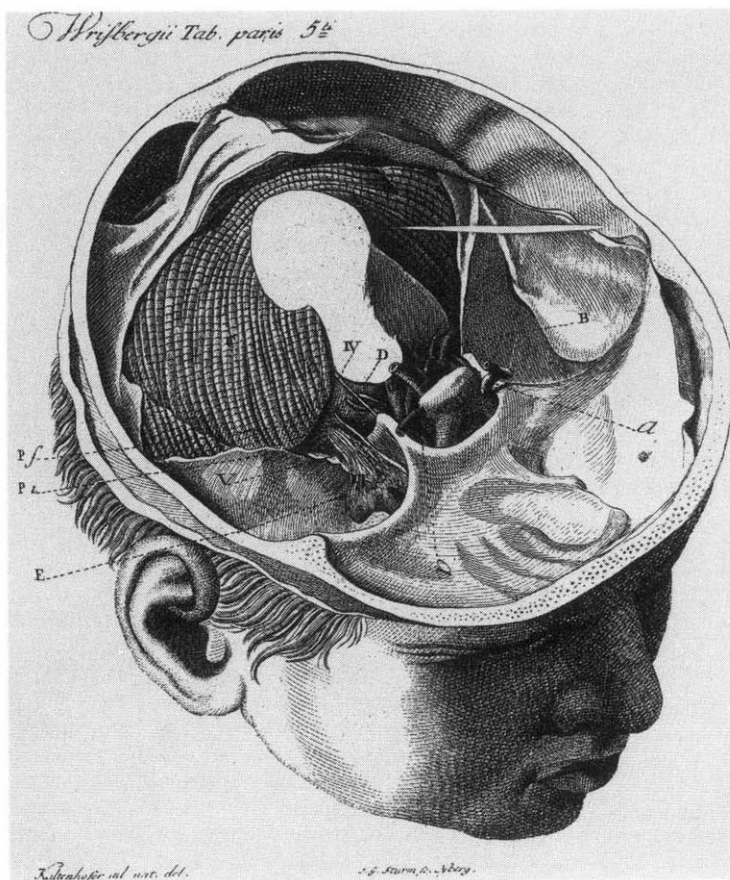
chapter. In some cases, important phenomena, such as lateral inhibition or neuromodulation, are dealt with sketchily at the expense of elaborating a particular technique or spelling out specific results.

Artificial intelligence is a particular area surveyed in relative detail. We read how connections between neurons, both real and simulated, were originally accorded a degree of preprogramming of 'significance'. These models were superseded by the more appealing 'neural darwinist' model of Gerald Edelman, which concerns the dynamic and empirical neural categorization of incoming information. But the reader is left with the impression not only that neural darwinism can overcome all

the difficulties of earlier models, but also that it can account for individual differences and even contribute radically towards our understanding of the brain. Such claims should surely be tempered by our imperfect knowledge of events at two extreme levels: the individual junction between single cells, and the aggregation of cells as mind-brain. A theoretical mechanism whereby enabling junctions are able to form labile and sensitive neuronal groups, which subsequently might collectively determine variable 'darwinistic' net responses, may well find a physiological counterpart in phenomena such as long-term potentiation. But surely the theoretical model incorporating enabling junctions should accommodate the possibility that the physicochemical responses generated at this building-block level need not necessarily constitute digital switching, that is, 'neural skinner-

ism'. Rather, neuronal responses may be analogue (parasympathetic transmission; dendritic versus axonal release), qualitatively diverse (intermittent co-release of neuroactive substances), and at least partially unidentified: these properties, which are not mentioned at all in the book, would lead in truly physiological situations to graded, possibly modulatory influences on neighbouring neurons that would constitute an unpredictable yet deterministic (that is, chaotic) effect.

At the other end of the spectrum, the physical basis of the conscious mind hardly crops up at all as a relevant issue,



Nervous beginnings: Heinrich August Wrisberg (1739–1808) erroneously denied with this sketch that branches of the trigeminal nerve run through the dura mater, despite proofs to the contrary that he himself had collected.

scanning. The range of subjects, though wide, is far from balanced. Although it would be in the nature of a general readership not to expect an exhaustive survey, the complete omission of any discussion of motor control at the expense of two chapters on vision cannot fail to give a distorted picture. The theme that glues this somewhat disparate material together is, according to the preface, the tension between the complementary views of functional localization and distributed plasticity. But this framework becomes obscured by both the thesis and detail of each individual

in contrast to its extensive coverage in the second section. Oliver Sacks attempts to redress the balance in an epilogue emphasizing the importance of the integrity of the individual as therapy in brain disorders and invokes neural darwinism as a *deus ex machina*. But this holistic approach floats free of any real neurophysiological infrastructure and in any case is simply too little too late.

In general, however, this third section is of value in that it faithfully reflects the limitations of current research on the brain. There is a tendency to smugness among neuroscientists, engendered in part by the prevalence of technocratic attitudes and exemplified in this section, that they are set fair towards discovering how the brain works. But surely we neuroscientists, who are all so increasingly defined by, and dependent on, technique, must first become fluent in the abstracted concepts and questions that preoccupied our predecessors, and that ultimately now underscore our experimentation. *The Enchanted Loom* offers a unique reference source for neuroscientists to benefit from their sherringtonian roots, and allows the general reader to savour the visual beauty derived from exploring the brain. □

Susan A. Greenfield is in the Department of Pharmacology, University of Oxford, Oxford OX1 3QT, UK.

New in paperback

■ Ellis Horwood have recently published *Neuroscience: An Illustrated Guide* by Roger A. Barker which aims to provide an overview and guide to current knowledge about the nervous system. Price is £19.95, \$39.95.

■ Two books on computing are issued in paperback by Oxford University Press. *The Universal Turing Machine: A Half-Century Survey* edited by Rolf Herken was described by Stephen Muggleton as "a vital reference work for computer scientists and mathematicians" (*Nature* **335**, 597; 1988). Price is £25. For students and professionals, as well as nonspecialists, is issued the third edition of *Dictionary of Computing* edited by Valerie Illingworth. Price is £6.99, \$10.95.

■ *The Frozen Earth: Fundamentals of Geocryology* by Peter J. Williams and Michael W. Smith is the first book in English to concentrate on the effects of cold climates on the surface of the Earth. Published by Cambridge University Press, price £19.50, \$29.95.

■ For those interested in quantum mechanics comes *Quantum Implications: Essays in Honour of David Bohm* edited by B. J. Hiley and F. David Peat. The collection was praised by Tony Sudbery as being "a fitting tribute to one of the most searching thinkers in modern physics" (*Nature* **331**, 26; 1988). Published by Routledge, price £9.99.

■ Finally, to mark the birth of Michael Faraday 200 years ago this week, Adam Hilger have issued *Michael Faraday and the Royal Institution* by J. M. Thomas. The book contains many drawings and photographs published for the first time. Price is £12.95. □

Nuclear citizens

Alan Cottrell

Nuclear Choices: A Citizen's Guide To Nuclear Technology. By Richard Wolfson. MIT Press: 1991. Pp. 467. \$33.75, £22.50.

Meltdown: The Collapse of the Nuclear Dream. By Crispin Aubrey. Collins & Brown: 1991. Pp. 190. Pbk £6.99.

The Nuclear Lion: What Every Citizen Should Know About Nuclear Power and Nuclear War. By John Jagger. Plenum: 1991. Pp. 402. \$29.94.

CITIZENS have opinions and the right to express them. They also have, less certainly, heads for thinking, as well as hearts for feeling. The nuclear age has presented new arenas for the exercise of these faculties. Although all three of the above books recognize this — indeed, it provides their common guiding motivation — they could hardly be more different in their approaches.

At the cerebral extreme is Wolfson's *Nuclear Choices*. The author follows the conventional scientist's approach: people have only to be given all the relevant basic scientific facts, clearly and without bias, and all will be well. Sound opinions will then be formed out of which sensible answers can always be constructed, by do-it-yourself means, to meet any situation. In his own words, "I have written this book on the premise that nuclear choices are best made by citizens who know something of the underlying issues, who understand the basis of nuclear technology, and who can judge for themselves statements advocating particular positions". The book is thus replete with scientific facts and explanations, skilfully presented somewhat in the style of *Scientific American*, but at a generally more elementary level. Although intended for people "with no particular scientific, technological or political background", the book would undoubtedly give the assiduous general reader a formidable knowledge with which to tackle the great issues of the nuclear age: nuclear warfare, radioactivity in the environment, Chernobyl and Three Mile Island, nuclear waste, benefits and risks of nuclear medicine and so on.

True to his scientific attitude, the author is entirely objective and impartial in his approach: rather like a judge addressing the jury at the end of a long case, summing up all the facts and pointing to the important questions to be answered, but refusing entirely to lead with his own views.

All of which is in total contrast to Aubrey's *Meltdown*. Here it is all heart, and a passionate anti-nuclear one at

that. Everything nuclear becomes a horror story, presented in lurid language. Nuclear sites are notorious; Britain is a nuclear dustbin; accidents at nuclear plants are invariably serious; nuclear plans are great white hopes, their projects white elephants; relations between government bodies and the nuclear industry are of course cosy; and independent pro-nuclear people are crazy.

There are few scientific facts in this book, which is perhaps as well, since those that there are are pretty inaccurate. For example, the moderator in a nuclear reactor "moderates the nuclear reaction by literally slowing it down"; radiation causes "genetic damage to a developing foetus"; and, "as the Windscale fire had shown, once the nuclear reaction had got out of control . . .". So much for the facts. But then, what do they matter to a heart in full frenzy, which already knows all the answers by instinctive conviction and is hermetically sealed in its certitudes? Let not the native hue of resolution be sicklied o'er with the pale cast of thought.

The most interesting part of the book is the description of the events surrounding the public enquiry into the proposed Hinkley Point C nuclear power station; and the associated 'stop Hinkley expansion' campaign, of which the author was one of the two full-time organizers. Provided that one is prepared to read this through the unavoidable antinuclear spectacles, it is an engaging account of a small historical episode in the nuclear age.

Jagger's *The Nuclear Lion* bases its title on a child's symbol of unbridled, mysterious and fearsome power: "to the general public, nuclear energy is a lion. It is a fearsome and terrible thing. And . . . this fear has arisen because it is not understood". And so, like Wolfson, he sets out to convert this fear into rational opinion by providing the necessary understanding. His presentation of the scientific facts is similar to Wolfson's, but not as detailed. Where his book differs strikingly however, and in this he takes up a position between Wolfson and Aubrey, is in its allowing both the heart and head full and equal expression.

Jagger grips the reader by the intensity of his passion, particularly that about the horrors of nuclear warfare. Writing obviously before both the recent US-Soviet agreement of strategic arms reduction and the equally recent discovery of the secret Iraqi military nuclear efforts, he finds a paradox in the fact that nations are "arming themselves to the teeth with apocalyptic nuclear weapons" yet have virtually stopped civil nuclear power because of fear of it. He is as strongly in favour of the latter, because it can easily be made very safe

and can give the world the energy it will need in the coming century, as he is against nuclear weapons and their military-industrial complex. It must seem incredible to Aubrey, to whom everything nuclear is unreservedly devilish, that anyone could passionately hold such diametrically opposite views about the civil and military aspects of nuclear energy. Yet many of us do. □

Sir Alan Cottrell is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

■ Also published recently on the politics of this subject: *The International Politics of Nuclear Waste* by Andrew Blowers, David Lowry and Barry D. Solomon, published by Macmillan, price £45 (hbk), £17.50 (pbk); *The Politics of Nuclear Power: A History of the Shoreham Nuclear Power Plant* by David P. McCaffrey, published by Kluwer, price Dfl. 150, \$89, £53. □

Out of Darwin's shadow

W. F. Bynum

Alfred Russel Wallace: An Anthology of His Shorter Writings. By Alfred Russel Wallace. Edited by Charles H. Smith. Oxford University Press: 1991. Pp. 551. £40, \$79.

ALFRED Russel Wallace (1823–1913) has always been a bit of an enigma. Even his contemporaries did not know quite what to make of a man who could with equal passion advocate at different periods of his life agnostic naturalism and committed spiritualism; who kept the Royal Society at a distance yet accepted two of its most prestigious medals; who presented himself as the mildest of men but seemed to thrive on public controversy; who was self-effacing even in his fat, double-deckered autobiography; who was the co-discoverer with Charles Darwin of the principle of natural selection and one of its chief exponents for three decades after Darwin's death, yet was content for 'darwinism' to be the sole descriptive appellation. For Darwin's part, he came to believe that he might have more to fear from friends like Wallace than announced adversaries; and the socially conventional Darwin did not live to see Wallace's most ardent espousal of socialism, land nationalization and phrenology, or his assault on smallpox vaccination, which he maintained was a scientific delusion.

While Darwin scholarship has become a mini-industry, with the publication of his correspondence and a new biography every few months, Wallace remains too



Illustration plate of mollucan beetles from Alfred Russel Wallace's *The Malay Archipelago*.

much in the shade. For, as Charles Smith's splendid anthology reveals, Wallace was a wonderfully creative thinker whose writings are still fresh a century later.

Wallace wrote easily, well and much. Smith's bibliography of Wallace's printed works lists almost 800 items, including 20 books, ranging from his classics *The Malay Archipelago* (1869) and *Island Life* (1880) to less familiar titles such as *On Miracles and Modern Spiritualism* (1875) and *Is Mars Habitable?* (1908). Wallace's last book was published in the year of his death at the age of 90. His hundreds of articles appeared in the mainline scientific press (proceedings and journals of the Linnean Society, Zoological Society, Entomological Society, and so on), high-brow Victorian periodicals (*Quarterly Review*, *Macmillan's Magazine*, *Fortnightly Review*, *Nineteenth Century*), and somewhat less conspicuous publications such as *The Eagle and the Serpent*, *The Medium and Daybreak* and *The Journal of Astrology*. To *Nature* he contributed a scientific paper, five letters to the editor (Sir Norman Lockyer) and two book reviews in the first volume (1869–70), and a letter to the same editor in the seventy-sixth volume 47 years later. Many issues in between contained something from Wallace's pen.

Smith's bibliography is by far the most complete ever compiled and the care that has gone into its preparation is reflected throughout the anthology. He has chosen more than 100 separate publications, ranging across the vast breadth of Wallace's interests, and varying in length from one to more than a dozen pages. Editorial intrusions are mostly short but to the point. The selection is really a guide to Wallace's thought, because he often subsequently incorpo-

rated into his books material first published in his articles.

That thought was always clear, whether Wallace was creatively exploring the nuances of biogeography or discoursing on the reality of the unseen world. At all times he considered himself to be a student of the experiential Universe, using the methods and tools of science to make sense of it all. The same tools were appropriate because evolution pervades the Universe, linking the past with the present, and joining the cosmic to the biological, social and psychological. As Smith comments, to Wallace social change was a manifestation of evolution no less than was the biological process sustained by natural selection.

Smith makes no attempt to whitewash Wallace by emphasizing his work on geographical distribution of plants and animals, mimicry, adaptation and ecology at the expense of the iconoclastic portions of his *oeuvre*. Indeed, the section on spiritualism is almost as long as that on biogeography, although Wallace turns up as a geographer in other sections as well. A concern with spatial relationships underlay much of his best scientific work.

It would be too much to suggest that the democratic, self-taught, lower-middle-class Wallace needs rehabilitation in an age like our own. Nevertheless, he deserves his own ground outside Darwin's shadow, which is where Smith has firmly placed him. Among other things, this volume reminds us that the letter that Wallace sent to Darwin in 1858 from half-way across the world was but one episode in Wallace's long and eventful life. □

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

Classified information

M. J. Crawley

British Plant Communities. Vol. 1. Woodlands and Scrub. Editor-in-chief J. S. Rodwell. Cambridge University Press: 1991. Pp. 395. £70, \$150.

It is curious that in a place like Great Britain, where there is such an outstanding tradition of mapping the detailed distribution of plant species, it should have taken so long to produce a comprehensive classification of our plant communities. This probably reflects two traits peculiar to British ecologists: on the one hand, a pragmatic admission that classification of woodlands at an everyday level is trivially easy (for example, oak woodland versus beech woodland); on the other, a more deep-seated feeling that distributions within woodlands are controlled by factors so numerous, complex and interacting that any attempt at a more elaborate taxonomy of plant communities would be futile. Thus, while most British plant ecologists are happy to divide oak woodlands into bluebell woods and bracken woods, and to split beech woodlands into dogsmerry woods and bramble woods, many feel uneasy with the full-blown phytosociological approach adopted by some of their continental colleagues.

This book steers a middle course between the extremes of oversimplification and undue elaboration. Its aim was to provide a simple easy-to-use list of plant communities that could be adopted by the Nature Conservancy Council in the production of vegetation maps. In this respect it has been highly successful, the draft keys and woodland descriptions having already been praised for their effectiveness.

Subjective assessment

The philosophy behind the classification was to 'let the plants do the talking', and so the communities are defined on the basis of floristic similarity without any recourse to environmental factors such as geographical location, substrate and altitude. But the study plots may well have been selected on the basis of prior conceptions about the range and composition of different woodland types, meaning that the resulting list of communities sometimes has an air of self-fulfilling prophecy about it. Nevertheless, the degree of selection is much less blatant than in earlier studies of British vegetation, such as that in *A Nature Conservation Review* (Cambridge University Press, 1977). The fragments of woodland left in Britain are so small and

so heavily influenced by their histories of management, that almost any random sample area of 50 × 50 metres is likely to be internally heterogeneous or to contain 'edge effects' of one sort or another. This is the principal justification for locating the study plots in subjectively assessed 'representative' stands rather than randomly.

The scheme recognizes 19 woodland types — 4 oak, 3 beech, 3 willow, 3 alder, 2 ash and 1 each of birch, pine, yew and juniper. Each of these is divided into subcommunities, typically on the basis of the ground flora. There are species lists and dot-distribution maps for each subcommunity type. For each community type there is an annotated list of synonymy, lists of the constant and rare species, a detailed discussion of the physiognomy of the woodland (including variations in the canopy, ground layer and shrubs of the understorey), descriptions of the subcommunities, a full account of habitat, notes on zonation and succession, details of geographical distribution, and mention of affinities with woodlands in other parts of the northern temperate region (including comparisons with European phytosociological classifications). The index of plant species is excellent, listing all the woodland and scrub community types in which a given plant is found; constant species appear in bold type, and species that are constant in one or more subcommunities are in italics. There is a comprehensive bibliography.

I tested the keys on a pedunculate oak woodland fringing the estuary of the River Fowey in southern Cornwall, and on the species-rich scrub on the chalk under-cliff at Folkstone Warren. In both of the cases the key worked extremely well; the oak woodland turned out to be W10 (*Quercus robur*-*Pteridium aquilinum*-*Rubus fruticosus* woodland; *Hedera helix* subcommunity), the scrub to be W21 (*Crataegus monogyna*-*Hedera helix* scrub; *Viburnum lantana* subcommunity). It is one thing to name a community, but much more difficult to predict its species composition. Again, the book worked well: all but 8 of the 67 species found in the woodland were predicted by the classification.

A difficulty with the presence/absence approach to classification is that it ignores variation in the relative abundance of the dominant plant species. Thus, two woodlands with the same classification can look completely different. For example, woodlands belonging to W10 can have a dense shrub layer or no shrubs at all, and a completely bare to a completely vegetated ground layer.

No one disputes that there are characteristic assemblages of species worthy of the name 'plant community', but it is a moot point whether the notion of defin-

able communities has added anything of substance to our understanding of vegetation. The description of the fine-scale variation within plant communities is likely to tell us more about the history of small-scale disturbances, historical contingencies and local differences in land management than about the tendency of pre-adapted species to group in particular ways. Instead, might there be characteristic combinations of environmental conditions (such as latitude, altitude, aspect, soil type, substrate stability, drainage) that define the subset of plant species that can thrive at a given site? The group of plants growing in the site is then nothing more than a sample from a pool of available species, its composition being influenced as much by chance as by biological interactions between plant species. Indeed, the extent to which such interactions determine the structure of a plant community and the resulting systemic integrity of the group of species are both highly contentious issues. The editors of the book are aware of these difficulties and are self-effacing about their efforts, having "tried to be honest about admitting deficiencies of coverage and recognising much unexplained floristic variation".

More than a list

Any quibbles about omissions (there is no mention of aspen woods or of urban *Buddleja* scrub) are outweighed by the excellence of the material included in the volume. Given the current interest in vegetation change under global warming, I would like to have seen more about alien trees such as *Quercus cerris* and *Robinia pseudoacacia* (although commoner aliens such as sycamore and sweet chestnut get reasonable coverage).

This is a splendid work, sure to provide the basis for classifying woodland plant communities for decades to come. It does much more than simply list the species of different woodland communities, and provides a wealth of information on the consequences of various kinds of woodland management, on the effects of habitat and on geographical differences in the composition of ground flora. Much of this is distilled from the vast experience of the editorial team and is published for the first time.

The book will provide ecologists with a nearly inexhaustible supply of questions about community structure and function, which should stimulate a new phase of field experimentation and theoretical modelling. It is a pity that there is no prospect of a zoological equivalent, even for our modest and relatively well-known fauna. □

M. J. Crawley is in the Department of Pure and Applied Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK.

Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Palaeocene

J. P. Kennett* & L. D. Stott†

* Marine Science Institute and Department of Geological Sciences, University of California, Santa Barbara, California 93106-6150, USA

† Department of Geological Sciences, University of Southern California, Los Angeles, California 90089-0740, USA

A remarkable oxygen and carbon isotope excursion occurred in Antarctic waters near the end of the Palaeocene (~57.33 Myr ago), indicating rapid global warming and oceanographic changes that caused one of the largest deep-sea benthic extinctions of the past 90 million years. In contrast, the oceanic plankton were largely unaffected, implying a decoupling of the deep and shallow ecosystems. The data suggest that for a few thousand years, ocean circulation underwent fundamental changes producing a transient state that, although brief, had long-term effects on environmental and biotic evolution.

WE describe foraminiferal oxygen and carbon isotope changes over the Palaeocene/Eocene transition at high stratigraphic resolution in an Antarctic sedimentary sequence. We infer palaeoenvironmental changes that caused major deep-sea benthic faunal extinctions at the end of the Palaeocene, perhaps the largest during the last 90 Myr (refs 1, 2). This event profoundly affected oceanic benthic communities deeper than the continental shelf (>100 m; neritic zone)^{1–4} resulting in a 35–50% species reduction of benthic foraminiferal taxa^{4,5}. The extinction level pre-dates the last appearance of *Morozovella velascoensis* (at the tropical P6a/P6b boundary^{1,6,7}) and is close to other biostratigraphic levels at or close to the Palaeocene/Eocene boundary^{1,8}. It is also located within one of the largest negative $\delta^{13}\text{C}$ changes of the Cenozoic, beginning in the late Palaeocene at ~60 Myr and continuing into the early Eocene^{6,9–11}. The $\delta^{13}\text{C}$ values immediately preceding the event are the highest for the entire Cenozoic⁹. This event is also located within a long-term negative $\delta^{18}\text{O}$ trend of >1‰ (refs 6, 9, 12–15), interpreted¹⁶ to reflect a warming of Antarctic surface waters by ~5 °C.

What could have caused such extensive changes in the ocean deeper than the continental shelf, an ecosystem that forms >90% by volume of the Earth's total habitable environments¹⁷? Before now, the extinction event has been studied at insufficient stratigraphic resolution to determine its speed and potential causes. During our earlier stable isotope investigations^{12,16} of Antarctic Palaeogene sequences, we discovered a sudden, brief, striking negative isotope excursion in planktonic and benthic $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ coinciding with the major benthic foraminiferal extinctions⁵ close to the Palaeocene/Eocene boundary. Our subsequent detailed studies reveal correlations in the timing of a number of isotopic changes and the extinctions (Figs 1, 2), thus providing a stronger basis for evaluation of the possible environmental changes that caused them.

The Palaeocene/Eocene transition has been examined in Ocean Drilling Project (ODP) hole 690B (65° 09' S, water depth 2,914 m) on the flank of Maud Rise, Weddell Sea, Antarctica. Palaeodepths during the Palaeocene/Eocene transition were

~2,100 m (ref. 12). The Palaeogene sequence^{18,19} consists of a nannofossil ooze with well-preserved planktonic and benthic foraminifera. There is a minor terrigenous sedimentary fraction (5–15% clay and ~10–15% mica)¹⁸. Colour photographs of the cores indicate a lack of bioturbation for several centimetres (core 19-3, 75–65 cm) encompassing the extinction level, whereas underlying and overlying sediments were clearly bioturbated. The cores show no drilling disturbance.

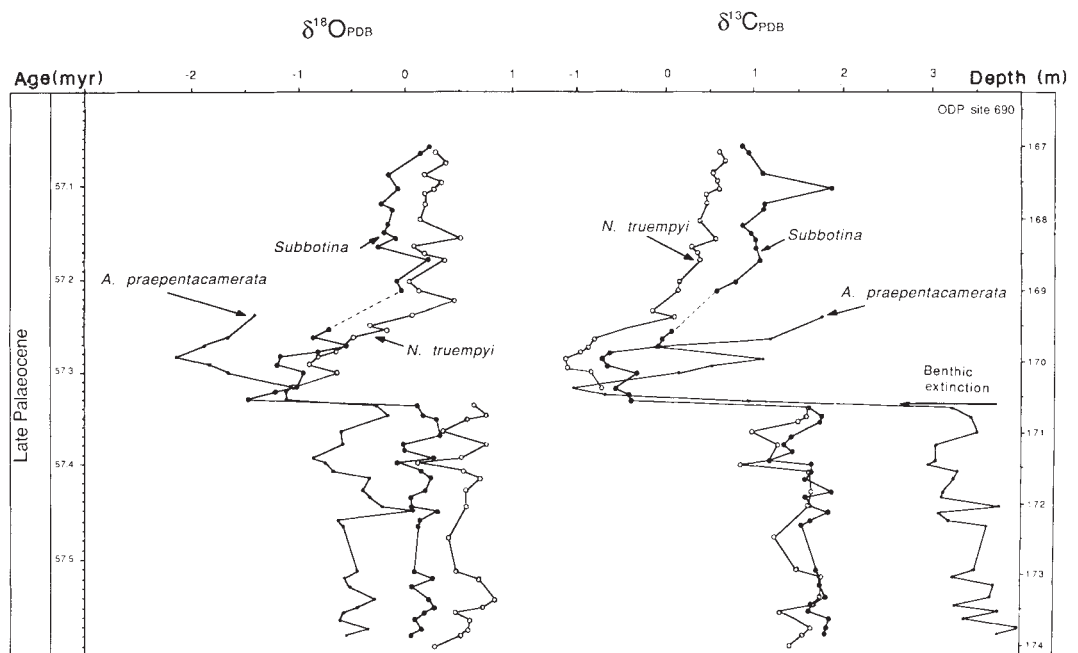
Chronology and approaches

The Palaeocene/Eocene boundary in hole 690B is located in a long reversed-polarity zone identified^{20,21} as Magnetochron 24R. As before²⁰, we place the boundary magnetostratigraphically at 57 Myr within Chron 24R^{21,22} at 166 m below sea floor (mbsf). This is close to the first appearance of *Planorotalites australiformis* (which defines the AP4/AP5 boundary at 170.05 mbsf; core 19-3, 15–18 cm) and below the first appearance of *Acarinina wilcoensis berggreni* (159.8 mbsf). The principal benthic extinction level (core 19-3, 70 cm) is ~40 kyr older, and is assigned an age of 57.33 Myr (Fig. 1). Our chronology is based on an age for the top of Chron 25N (185.2 mbsf) of 58.64 Myr and the bottom of Chron 24N (154.62 mbsf) of 56.14 Myr (ref. 21). This provides an average sedimentation rate of 1.23 cm per 1,000 yr (1 cm in ~800 yr). In low-latitude sequences, the Palaeocene/Eocene boundary has been correlated to the last appearance of *M. velascoensis*, marking the boundary between the P6a and P6b zones with an assigned age of 57.8 Myr (refs 6, 7, 22). More recent correlations²³ place the younger boundary at 57 Myr and within zone P6b, postdating the extinction of *M. velascoensis*. This species is not present in site 690 and the AP4/AP5 boundary remains to be accurately correlated with low-latitude biostratigraphy including the P6a/P6b boundary. We believe that the chief deep-sea extinction level near the end of the Palaeocene is synchronous in marine sequences, and that offsets between our age estimates (Fig. 1) and others^{6,22} result either from biostratigraphic miscorrelations or the necessarily broad age interpolations between magnetochrons at site 690.

Foraminiferal samples were analysed for oxygen and carbon isotope composition between 174.00 mbsf (57.59 Myr) and 166.95 mbsf (57.06 Myr) at a maximum sampling interval of 10 cm, a stratigraphic resolution of ~4–7.5 kyr. We located the extinction horizon within a 4-cm interval between 170.62 mbsf (core 19-3, 72 cm) and 170.58 mbsf (core 19-3, 68 cm), estimated to represent ~3 kyr (57.334 and 57.331 Myr) (Fig. 2). This we sampled at 1-cm intervals, providing an age resolution of ~800 yr (Fig. 2).

Isotope analyses were conducted on monospecific samples of the planktonic foraminifers *Acarinina praepentacamerata*, *Subbotina patagonica* and *Subbotina varianta*, and the benthic foraminifer *Nuttallides truempyi*. This benthic species occurs throughout the record, except for a brief gap close to the extinction event. We used standard approaches to extract and prepare the foraminifera for faunal and isotopic analyses¹². Oxygen isotopic palaeotemperatures were determined assuming an ice-free world (mean ocean water $\delta^{18}\text{O}$ = -1.2‰ relative to present day¹³).

FIG. 1 Changes in $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ ($[(^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{standard}} - 1] \times 1,000$; $[(^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{standard}} - 1] \times 1,000$) of the planktonic foraminifers *Acarinina praepentacamerata* and *Subbotina* and the benthic foraminifer *Nuttallides truempyi* in the latest Palaeocene (57.0 to 57.6 Myr) of Ocean Drilling Program (ODP) hole 690B, Maud Rise, Antarctica. Note the sharp negative shifts at 57.33 Myr coinciding with the main level of extinction in benthic foraminifera. Oxygen isotope values have been adjusted to account for non-equilibrium effects^{6,12}. The gap in isotope records of *Subbotina* is between the last appearance of *S. patagonica* and the first appearance of *S. varianta*. The upper extent of isotope records of *A. praepentacamerata* marks the disappearance of this species. Age framework is after ref. 23. The age assignments will change, however, because the age of the Palaeocene/Eocene boundary is being revised to 2 Myr younger⁵⁶. Depth scale at right represents metres in section below sea floor. Samples were analysed with a Finnigan MAT 251 mass spectrometer linked to a Carousel-48 automatic carbonate preparation



device. The laboratory gas standard is calibrated to PDB by cross-calibration using the standard NBS 20. Precision is 0.1‰ or better for both $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$. We applied the standard $\delta^{18}\text{O}$ correction factor of +0.5‰ to *Nuttallides truempyi*^{6,12}. No correction factor was made to $\delta^{13}\text{C}$ for this taxon.

The terminal Palaeocene isotopic excursion

The planktonic foraminifer *A. praepentacamerata* records the lowest oxygen and the highest carbon isotope values (Fig. 1). Such values are consistent with a near-surface habitat, possibly during the austral spring/summer months^{12,16}. The species of *Subbotina* record higher $\delta^{18}\text{O}$ and lower $\delta^{13}\text{C}$ values (Fig. 1) than *A. praepentacamerata*, indicating a deeper-water planktonic habit and/or a preference for cooler months of the year¹⁶. As expected in a thermally stratified ocean with deep waters richer in nutrients, the highest $\delta^{18}\text{O}$ and lowest $\delta^{13}\text{C}$ values are shown by the benthic foraminifer *N. truempyi*.

Between 57.6 and 57.33 Myr, isotope values showed relatively little change (Fig. 1). Surface water isotopic temperatures are estimated to have been 13–14 °C and deep water temperatures ~10 °C. The oxygen and carbon isotopic composition of planktonic and benthic foraminifera (Fig. 1) show a considerable excursion between ~57.33 and 57.22 Myr. This excursion (the 'terminal Palaeocene isotopic excursion') began abruptly at ~57.33 Myr with a conspicuous decrease in $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values, followed by a return to values only slightly lower than before the excursion (Fig. 1).

Oxygen isotopic changes. Between ~57.33 and 57.32 Myr, the marked decrease in $\delta^{18}\text{O}$ occurred in all three foraminiferal groups. The change was largest (2.0‰) in the benthic forms, intermediate (1.5‰) in deeper-dwelling planktonics and smallest (1.0‰) in shallow-dwelling planktonics. Surface water temperatures increased to ~18 °C. This isotopic shift coincided with the benthic extinctions (Figs 1 and 2). Isotope values in shallow surface water further decreased to ~-2.2‰, suggesting temperatures of ~21 °C, between 57.31 and 57.28 Myr, probably the warmest of the Cenozoic¹⁶. In contrast, the isotope temperatures of *Subbotina* decreased after 57.32 Myr. The rapid warming is associated with an increase in warm subtropical planktonic microfossils in Antarctic waters, including discoasters²⁴ and morozovellids. Also, an associated peak in kaolinite suggests warm, humid climate on Antarctica, the inferred source of the clays²⁵.

A brief gap (20 kyr) exists in the range of *N. truempyi* during

the excursion (Fig. 1). When *N. truempyi* reappeared at ~57.31 Myr, its oxygen isotope values were similar (-1.1‰) to the planktonic forms (Fig. 1). Thus, deep waters warmed to temperatures close to those at the ocean surface, temporarily eliminating the vertical temperature gradient at this location. This gradient was re-established ~30 kyr after the initial isotope decrease (Fig. 1). Following the warming peak, temperatures decreased, but remained ~1–2 °C higher than before the beginning of the excursion (Fig. 1).

Carbon isotopic changes. The pattern of $\delta^{13}\text{C}$ change during the excursion (Fig. 1) was similar to that of $\delta^{18}\text{O}$: a marked initial decrease in values at ~57.33 Myr followed by a more gradual increase. The change was very large (>4‰) in the surface-dwelling planktonic *A. praepentacamerata*. *Subbotina* showed a 2.0‰ decrease. The absence of benthic values during the shift prevents a detailed comparison, but values rapidly decreased by at least 2.6‰, from 1.5‰ at 57.33 Myr to -1.1‰ by 57.29 Myr. From ~57.33 to 57.31 Myr, $\delta^{13}\text{C}$ values of surface and deep planktonics are similar; *Subbotina* values are between ~-0.5‰ and -0.8‰. Benthic $\delta^{13}\text{C}$ values at ~57.32 Myr are almost identical to surface-water planktonic values. The large $\delta^{13}\text{C}$ gradient (~2.0‰) that had existed between surface and deep waters before the excursion was virtually eliminated at ~57.32 Myr. After 57.30 Myr, surface-water $\delta^{13}\text{C}$ values began to increase again and a vertical gradient was re-established.

A decrease in benthic foraminiferal diversity (Fig. 2) coincided with the $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ changes in the deeper-dwelling planktonic *Subbotina* and the $\delta^{18}\text{O}$ changes in the shallow-dwelling planktonic *A. praepentacamerata*. Most of the reduction in benthic diversity had already occurred, however, before the initiation (~3 kyr later) of the principal $\delta^{13}\text{C}$ decrease in the surface-dwelling *A. praepentacamerata* (Fig. 2). Thus, the interval of rapid isotope change is not the result of a hiatus. Within ~3 kyr, surface water temperatures at hole 690B are estimated to have increased by 3–4 °C, with a total warming in subsurface waters of 6 °C in 6 kyr. The $\delta^{13}\text{C}$ shift (2.0‰) in *Subbotina* lasted 6 kyr. Most (3.5‰) of the $\delta^{13}\text{C}$ shift in *A. praepentacamerata*, which began slightly later (Fig. 2), occurred within 11 kyr,

although the trend continued for more than 20 kyr. After the shift, the vertical $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ gradients were temporarily eliminated or even reversed.

Changes in carbon isotope gradients ($\Delta\delta^{13}\text{C}$) between shallow and deep forms (Fig. 2) indicate that characteristics of the vertical water column differed before, during and after the isotope excursion. A distinct $\delta^{13}\text{C}$ gradient between surface and deep waters existed before 57.33 Myr. Furthermore, at that time the similarity between the $\delta^{13}\text{C}$ composition of the benthic foraminifera and the subbotinids suggests the existence of a midwater zone with a strong minimum in oxygen concentration. During the $\delta^{13}\text{C}$ shift beginning at ~ 57.33 Myr, the gradient between all groups was eliminated (Fig. 2). After 57.29 Myr, the vertical $\delta^{13}\text{C}$ gradient was gradually re-established (Fig. 1). But after re-establishment, by ~ 57.28 Myr, a distinct isotope offset existed between benthic foraminifera and the subbotinids, suggesting a reduction in the strength of the oxygen minimum zone.

Biotic changes

Large changes in the taxonomic composition of benthic foraminifera clearly define the extinction event and the pattern of decreasing diversity⁵. Before the extinction event, the simple diversity of benthic foraminiferal assemblages ($>150\ \mu\text{m}$) averaged ~ 55 – 60 species (Fig. 2) or higher⁷. Assemblages containing abundant benthic foraminifera include a wide range of morphotypes inferred to be of both infaunal and epifaunal habit^{5,26}, including trochospiral and other coiled forms. A considerable decrease in diversity ($>150\ \mu\text{m}$) coincides with the $\delta^{18}\text{O}$ shift (Fig. 2). Simple diversity dropped by 72% from ~ 60 to 17 species ($>150\ \mu\text{m}$) in <4 cm (<3 kyr). Many distinctive taxa such as *Stensioina beccariiformis* and *Neoflabellina* disappeared early and rapidly (<1.5 kyr) during the isotope shift marking the extinction event (Fig. 2). Most trochospiral forms such as *N. truempyi* had disappeared by the midpoint of the shift. Most lost taxa are not observed again even in the immediately overlying 1-cm sample interval, supporting sedimentary and isotope evidence for a lack of bioturbation over this interval. Benthic foraminifera ($>150\ \mu\text{m}$) also became uncommon to rare (as low as ~ 30 specimens per sample) during the extinction event. Small specimens ($<150\ \mu\text{m}$), however, remained abundant throughout this interval. Ostracods decreased markedly in diversity and abundance during the shift, leaving a rare assemblage composed almost exclusively of small, smooth, thin-walled forms. Both planktonic and benthic foraminiferal assemblages are well preserved throughout, showing no apparent increase in dissolution. During this episode, planktonic foraminifera and calcareous nannofossils, which remained abundant, increased in diversity in hole 690B (refs 20, 24) and elsewhere in the oceans²⁷, possibly as a result of ocean surface warming.

The 5-kyr interval immediately following the isotope shift (Fig. 2) is marked by low diversities and abundances of benthic foraminifera ($>150\ \mu\text{m}$ fraction). Coiled forms are almost completely absent, leaving an assemblage dominated by relatively small and thin-walled uniserial, triserial and other forms typical of an infaunal habitat^{5,26}. The relative increase in abundances of small specimens of benthic foraminifera associated with a decrease in diversity suggests conditions low in oxygen and high in nutrients^{5,28}. The diversity of benthic foraminiferal assemblages ($>150\ \mu\text{m}$) increased thereafter (Fig. 2) to an average of ~ 30 species ($>150\ \mu\text{m}$). This increase resulted, in part, from the reappearance of forms previously present, but a high proportion of species ($\sim 35\%$) in hole 690B became truly extinct at the beginning of the excursion⁵.

Major, roughly coeval extinctions of benthic foraminifera have been recorded in other deep-sea sequences^{1,7,29,30} and in upper bathyal marine sequences in Trinidad³, the Reichenhall and Salzburg Basins, Austria³¹ and northern Italy³². In contrast, benthic foraminiferal assemblages on the continental shelves were little affected. The loss of so many characteristic Late Cretaceous and Palaeocene deep-sea taxa produced a deep-sea

benthic foraminiferal assemblage of Tertiary character¹. The biotic crisis at the K/T boundary was opposite, in that elements of the oceanic plankton and shallow benthic communities were considerably reduced, whereas deep-sea benthic foraminiferal assemblages experienced little change^{4,33–36}.

Interpretations of biotic changes

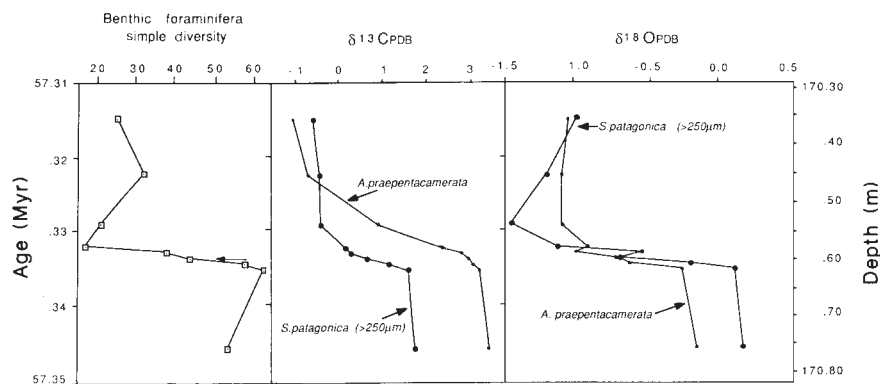
An important question relates to the large-scale character of the oceanographic changes that caused these rapid biotic changes. The lack of major extinctions in oceanic plankton and in shallow-water benthic communities eliminates any possibility that the changes were caused by a bolide impact with the Earth, as has been suggested for the terminal Cretaceous extinctions³⁴. A more plausible explanation would involve an oceanic cause. The extinctions were clearly driven by processes that preferentially affected the deep-sea biota below the thermocline or continental shelf. The process that caused the extinctions, although not instantaneous, was remarkably rapid (<3 kyr), and any proposed mechanism must have had the capacity to affect vast volumes of the deep ocean rapidly. The extinctions occurred at about the rate of the replacement time of the oceans (currently ~ 1 kyr (ref. 37), although possibly slower in the early Palaeogene). Also, the speed and magnitude of the associated temperature increase implies global, not solely oceanic, warming involving strong positive feedback mechanisms. The superposition of the rapid, negative $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ excursions upon similar, more gradual trends during the late Palaeocene to early Eocene^{13,4} suggests a threshold event similar to the oxygen isotope shift near the Eocene/Oligocene boundary²⁸, although in the opposite sense. Our data indicate that an essentially isothermal water column developed, with deep waters in the Antarctic warming more than shallow waters. The rapidity of the initial changes and the gradual return (over ~ 100 kyr) to isotope values similar to those before the event suggest an ocean–climate system temporarily in disequilibrium³⁷. As deep waters warmed, most benthic foraminiferal species ($>150\ \mu\text{m}$) were eliminated at this location. The disappearance of many more inferred epifaunal than infaunal benthic species ($>150\ \mu\text{m}$) indicates that species living in sediments were more likely to survive. Nevertheless, the disappearance of many infaunal forms indicates that the changes also affected the infaunal environment. Lack of bioturbation early in the excursion suggests that burrowing organisms were also severely affected.

Three main hypotheses have been proposed to account for the deep-sea benthic crisis at the end of the Palaeocene. They are: (1) a rapid warming of the deep ocean and a change in bottom water sources⁶; (2) a deep-sea oxygen deficiency due to the sudden warming and change in circulation of deep water^{5,11,12,19,36}; and (3) a sharp drop in surface ocean productivity that reduced trophic resources available for deep-sea benthic organisms^{10,39,40}.

The benthic extinctions occurred at the same time as rapid warming of deep waters and before the shift in $\delta^{13}\text{C}$ in the surface planktonic foraminifera. They therefore occurred before any associated large change in surface water processes that might have affected primary production. Although no large changes are evident in calcareous nannofossils at this location (A. R. Edwards, personal communication), elsewhere an earliest Eocene peak in taxonomic turnover was interpreted as a response to decreasing primary production⁴¹. Substantial productivity reduction is generally accompanied by a decrease in biogenic calcium carbonate accumulation, which is not observed at the Palaeocene/Eocene boundary at site 690. Considerable reduction in delivery of organic matter to the deep sea during the faunal crisis is unlikely^{4,11,12,36}, although it could have occurred later during the isotope excursion.

It is more likely that the benthic extinctions were caused by the rapid temperature increase of deep waters and associated reduction of oxygen concentrations^{4,16}. In the modern ocean,

FIG. 2 Changes in oxygen and carbon isotope composition of the planktonic foraminifers *A. praepentacamerata* (>63 μm fraction) and *Subbotina patagonica* (>250 μm fraction) and in simple diversity of benthic foraminiferal assemblages at high stratigraphic resolution during the negative isotope excursion in the latest Palaeocene shown at 57.33 in Fig. 1 (ODP hole 690B). Note the large, rapid decrease in diversity coinciding with the negative oxygen isotope shift, followed by a return to higher diversities. The large latest-Palaeocene extinction in benthic foraminifera coincides with the drop in diversity. Arrow at left indicates the last appearance of a distinctive Palaeocene benthic foraminifer *Stensiolina beccariiformis*. All benthic foraminifera (>150 μm) were picked from samples in the early part of the excursion to determine simple diversity. Because specimens were abundant before the excursion, simple diversity was determined by partial picking and thorough scanning of sample residues (>150 μm). No relationship is apparent



between the original dried weight of samples and the simple diversity and abundances of benthic foraminifera. Depth scale at right represents metres in section below sea floor.

cold (<6 °C) deep waters contain sufficient oxygen to oxidize organic matter fully and to maintain diverse benthic communities in most areas. At the same nutrient levels and Redfield ratios, however, deep waters of 10 °C would not contain sufficient preformed oxygen to avoid anoxia unless atmospheric oxygen levels were considerably higher or primary production lower than at present. Modelling studies suggest that atmospheric oxygen levels were possibly 10% higher in the early Palaeogene⁴², in which case deep waters of 10 °C could have remained aerobic if overall nutrient levels were no higher than in the modern ocean. But at temperatures higher than 15 °C, as at the Palaeocene/Eocene boundary, deep waters would almost certainly have been depleted in oxygen. Deep waters are unlikely to have become completely anoxic during the excursion because there is no increase in organic carbon accumulation at hole 690B and possibly elsewhere⁴³. Therefore a balance must have been maintained between deep-sea oxygen levels and the supply of organic carbon from surface waters and of preformed nutrients in deep waters⁴⁴. The decrease in $\delta^{13}\text{C}$ gradients suggests reduced primary production, contrary to benthic foraminiferal trends.

Interpretations of oceanographic changes

What process could have triggered the rapid oceanic warming and inferred reduction in deep-sea oxygen concentrations that produced the extinctions? It has been proposed^{4-6,11,12} that the rapid warming resulted from the concurrence of an almost total dominance in the oceans of warm saline deep waters produced at mid-latitudes, and a reduction of high-latitude deep water sources, in an extreme example of the Proteus ocean model¹². Modelling studies⁴⁵ suggest that deep-ocean oxygen concentrations are primarily controlled by deep-water formation processes. The relative importance of high- and mid-latitude sources of deep ocean waters during the early to middle Palaeogene is controversial (G. Mead and D. Hodell, manuscript in preparation; and refs 5, 6, 12). In the modern ocean, almost all dense, oxygen-rich waters are produced at high latitudes as a result of cold temperatures in combination with moderately high salinities³⁷. Saline, dense waters also form in the Mediterranean and Red Seas because of high net evaporation. Because of low buoyancy fluxes, however, they sink only to thermocline depths³⁷.

During the early Palaeogene, different global geography and climate⁴⁶⁻⁴⁸ combined to make the ocean circulation very different from that of the modern ocean^{12,46,49}. Significant Antarctic warmth^{6,19,20,22} coupled with inferred higher precipitation^{25,50-52} may have caused a large reduction in deep waters derived from the high latitudes^{6,12}. At the same time, the mid-latitude Tethys Seaway would have produced warm, saline deep waters¹², probably in sufficient volume⁴⁵ to alter the density distributions in the oceans considerably. We suggest that during

the excursion these factors combined to form, through positive feed-back responses, an extreme case of such an ocean, because a global climate threshold was temporarily penetrated. The nature of the triggering remains unknown.

The character of the $\delta^{13}\text{C}$ changes constrains hypotheses for the cause of the extinctions. Of particular importance is that the negative carbon isotope shift occurred throughout the water column (Fig. 2) and was smaller (~2.5‰) in deep waters than in surface waters (~4.0‰). This trend is opposite to that of the $\delta^{18}\text{O}$ record. Furthermore, the shift in $\delta^{13}\text{C}$ in waters at thermocline and greater depths coincided with warming of these waters, but most of the $\delta^{13}\text{C}$ shift occurred later in surface waters. A negative $\delta^{13}\text{C}$ shift of this magnitude³⁹ might be interpreted as indicating a transfer of light $\delta^{13}\text{C}$ organic carbon from the continents to the oceans^{43,53}, but this is unlikely because there is no evidence of any significant and rapid drop in global sea level at that time^{40,54}. Moreover, the diachronism shown in the carbon isotope changes between surface and deeper-dwelling planktonic forms does not support this hypothesis. It has been suggested⁴⁰ that plate-tectonic reorganization at this time increased oceanic volcanism and deep-sea hydrothermal activity, triggering warming of deep waters and global greenhouse warming by release of carbon dioxide. The speed of the palaeoenvironmental changes and brevity of the excursion lend little support to this hypothesis.

Summary

The rapid (<6,000 yr) negative oxygen and carbon isotope shifts in Antarctic waters near the end of the Palaeocene reflect oceanographic changes associated with global warming that caused considerable deep-sea extinctions. These observations show that for at least the early Cenozoic, short (1,000 yr) events can strongly affect the course of environmental and biotic evolution. Furthermore, such change can be effectively decoupled between different parts of the biosphere, in this case the deep and shallow marine ecosystems. The oceanographic changes associated with the excursion were clearly broad and complex, and seem to reflect a transient state between fundamentally different modes of ocean circulation. Brief elimination of the vertical $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ gradients indicates vertical ocean mixing and homogenization of nutrient distributions, presumably related to vertical instability of the Antarctic Ocean. The slightly delayed shift, during the excursion, to highly negative $\delta^{13}\text{C}$ values in Antarctic surface waters is difficult to explain but suggests higher concentrations of nutrients and total carbon dioxide, partial pressure of carbon dioxide⁵⁵. This would have contributed to greenhouse warming. Detailed examination of the excursion is required elsewhere, especially at low latitudes, for better definition of the character and magnitude of the oceanographic and biotic changes associated with this remarkable event. □

Received 13 February; accepted 11 August 1991.

1. Tjalsma, R. C. & Lohmann, G. P. *Micropaleont. spec. Publ.* **4** (Micropaleontology Press, New York, 1983).
2. Douglas, R. & Woodruff, R. in *The Oceanic Lithosphere, The Sea*, Vol. 7 (ed. Emiliani, C.), 1233–1327 (Wiley, New York, 1981).
3. Beckmann, J. P. *Reports Int. Geol. Cong. 21st Session, Norden*, Part V 57–69 (1960).
4. Thomas, E. in *Origins and Evolution of the Antarctic Biota* (ed. Crame, J. A.) *Spec. Publ.* No. 47, 283–296 (Geological Society, London, 1989).
5. Thomas, E. in *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) 571–594 (Ocean Drilling Program, College Station, Texas, 1990).
6. Miller, K. G. et al. *Paleoceanography* **2**, 741–761 (1987).
7. Berggren, W. A. & Miller, K. G. *Micropaleontology* **34**, 362–380 (1988).
8. Boltovskoy, E. & Boltovskoy, D. *Rev. Micropaleont.* **31**(2), 67–84 (1988).
9. Shackleton, N. J. et al. in *Init. Reports DSDP* **74** (eds Moore, T. C. et al.) 599–612 (US Government Printing Office, Washington DC, 1984).
10. Shackleton, N. J. *Palaeogr. Palaeoclimat. Palaeoecol.* **57**, 91–102 (1986).
11. Katz, M. E. & Miller, K. G. in *Proc. ODP, Scient. Results* **144** (eds Ciesielski, P. F. et al.) (Ocean Drilling Program, College Station, Texas, in the press).
12. Kennett, J. P. & Stott, L. D. in *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) 865–880 (Ocean Drilling Program, College Station, Texas, 1990).
13. Shackleton, N. J. & Kennett, J. P. in *Init. Reports DSDP* **29** (eds Kennett, J. P. et al.) 743–755 (US Government Printing Office, Washington DC, 1975).
14. Savin, S. M., Douglas, R. G. & Stehli, F. G. *Geol. Soc. Am. Bull.* **86**, 1499–1510 (1975).
15. Kennett, J. P. *Mar. Micropaleontol.* **3**, 301–345 (1978).
16. Stott, L. D., Kennett, J. P., Shackleton, N. J. & Corfield, R. M. in *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) 849–863 (Ocean Drilling Program, College Station, Texas, 1990).
17. Childress, J. J. in *Oceanography: The Present and Future* (ed. Brewer, P.) 127–135 (Springer, New York, 1983).
18. Barker, P. F. et al. *Proc. ODP, Init. Reports* **113** (Ocean Drilling Program, College Station, Texas, 1988).
19. Kennett, J. P. & Barker, P. F. in *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) 937–960 (Ocean Drilling Program, College Station, Texas, 1990).
20. Stott, L. D. & Kennett, J. P. in *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) 549–569 (Ocean Drilling Program, College Station, Texas, 1990).
21. Spiess, V. in *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) 261–315 (Ocean Drilling Program, College Station, Texas, 1990).
22. Berggren, W. A., Kent, D. V. & Flynn, J. J. *Mem. Geol. Soc. London* **10**, 141–195 (1985).
23. Aubry, M.-P. et al. *Paleoceanography* **3**, 707–742 (1988).
24. Pospichal, J. J. & Wise, S. W. Jr *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) 613–638 (Ocean Drilling Program, College Station, Texas, 1990).
25. Robert, C. & Mailliot, H. *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) 51–70 (Ocean Drilling Program, College Station, Texas, 1990).
26. Corliss, B. H. & Chen, C. *Geology* **16**, 716–719 (1988).
27. Aubry, M.-P., Gradstein, F. M. & Jansa, L. F. *Micropaleontology* **36**(2), 164–172 (1990).
28. Bernard, J. M. *J. Foram. Res.* **16**, 207–215 (1986).
29. Schnitker, D. in *Init. Reports DSDP* **48** (eds Montadert, L. et al.) 377–414 (US Government Printing Office, Washington DC, 1979).
30. Nomura, R. in *Proc. ODP, Scient. Results* **121** (eds Peirce, J. W. et al.) (Ocean Drilling Program, College Station, Texas, in the press).
31. von Hildebrandt, A. *Akad. Wiss. (Wien), Math.-Naturw. Klasse, Abh. n. ser.* **108**, 1–182 (1962).
32. Braga, G., De Biase, R., Grunig, A. & Proto Decima, F. in *Monogr. micropaleontol. sul Paleocene e l'Eocene di Possagno, Prov. di Treviso, Italia, Abh.-Mem. Suisse Paleontol.* (ed. Bolli, H. M.) **97**, 85–111 (1975).
33. Russell, D. A. *Ann. Rev. Earth planet. Sci.* **7**, 163–182 (1979).
34. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095–1108 (1980).
35. Stott, L. D. & Kennett, J. P. *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) 829–846 (Ocean Drilling Program, College Station, Texas, 1990).
36. Thomas, E. in *Geol. Soc. Am. spec. Publ.* **247** (in the press).
37. Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Eldigio, Palisades, 1982).
38. Kennett, J. P. & Shackleton, N. J. *Nature* **260**, 513–515 (1976).
39. Shackleton, N. J., Corfield, R. M. & Hall, M. A. *J. Foram. Res.* **15**, 321–336 (1985).
40. Rea, D. K., Zachos, J. C., Owen, R. M. & Gingerich, P. D. *Palaeogr. Palaeoclimat. Palaeoecol.* **79**, 117–128 (1990).
41. Corfield, R. M. & Shackleton, N. J. *Historical Biology* **1**, 323–343 (1988).
42. Berner, R. A. & Canfield, D. E. *Am. J. Science* **289**, 333–361 (1989).
43. Shackleton, N. J. in *Geol. Soc. Spec. Publ.* **26**, 423–434 (1987).
44. Herbert, T. D. & Sarmiento, J. L. *Geology* **19**, 702–705 (1991).
45. Brass, G. W., Southam, J. R. & Peterson, W. H. *Nature* **296**, 620–623 (1982).
46. Kennett, J. P. *J. geophys. Res.* **82**, 3843–3860 (1977).
47. Haq, B. U. *Oceanologica Acta* **4** (Suppl. to Vol. 4, Proc. 26th Int. Geol. Congress) 71–82 (1981).
48. Hay, W. W. *Geol. Soc. Am. Bull.* **100**, 1934–1956 (1988).
49. Benson, R. H. *Historical Biogeography, Plate Tectonics, and the Changing Environment* (eds Gray, J. A. & Boucot, A. J.) 379–389 (Oregon State University Press, 1979).
50. Truswell, E. M. *New Zealand Dept. Sci. Ind. Res. Bull.* **237**, 131–134 (1986).
51. Mohr, B. A. R. *Proc. ODP, Scient. Results* **113** (eds Barker, P. F. et al.) (Ocean Drilling Program, College Station, Texas, 1990).
52. Case, J. A. *Geol. Soc. Am. Mem.* **169**, 523–530 (1988).
53. Miller, K. G. & Fairbanks, R. G. *The Carbon Cycle and Atmospheric CO₂: Natural Variations, Archean to Present* (eds Sundquist, E. & Broecker, W. S.) *Am. geophys. Un. Monogr. Series* **32**, 469–486 (1985).
54. Haq, B. U., Hardenbol, J. & Vail, P. R. *Science* **235**, 1156–1166 (1987).
55. Sarmiento, J. L. & Toggweiler, J. R. *Nature* **308**, 621–624 (1984).
56. Berggren, W. A., Kent, D. V., Obradovich, J. D. & Swisher, C. C. in *Eocene-Oligocene Climatic and Biotic Evolution* (eds Prothero, D. R. & Berggren, W. A.) (Princeton University Press, in the press).

ACKNOWLEDGEMENTS. We thank D. Lea, B. Flower, and the reviewers, E. Thomas, K. Miller and J. Zachos, for constructive criticism of the manuscript, and M. Arthur for discussions. We also thank A. Puddicombe and H. Berg for technical assistance. This work was supported by the NSF Division of Polar Programs (J.K.) and Submarine Geology and Geophysics (J.K. and L.S.).

Primary structure and expression of bovine poly(A) polymerase

Tobias Raabe, F. J. Bollum* & James L. Manley

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

* Department of Biochemistry, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814, USA

Poly(A) polymerase has a critical role in the synthesis of messenger RNA in eukaryotic cells. The isolation and characterization of complementary DNAs encoding bovine poly(A) polymerase is described here. The predicted sequences of the mRNA and protein reveal features that provide insights into how the enzyme functions and how it might be regulated. Poly(A) polymerase expressed from a cloned cDNA is fully functional in *in vitro* assays, and mutational analyses have identified a putative regulatory domain that enhances, but is not essential for, activity.

AN enzyme activity able to polymerize AMP residues onto an RNA primer was first described over 30 years ago¹. Subsequent studies have described the purification and characterization of poly(A) polymerases from many different organisms (reviewed in refs 2, 3). Although all these enzymes share a requirement for an RNA primer and are specific for ATP as substrate, significant differences in several properties, such as subcellular

localization during fractionation (nuclear versus cytoplasmic) and apparent relative molecular mass (~40–80,000; M_r , ~40–80K), were observed. Although interest in these enzymes was heightened by the discovery of poly(A) tails on mammalian mRNAs^{4–6}, the involvement of the characterized enzymes in polyadenylation of mRNA precursors was clouded by the almost complete lack of primer specificity displayed by purified poly(A) polymerases.

The discovery of the highly conserved AAUAAA sequence in mRNA (ref. 7), the demonstration that this sequence is critical for polyadenylation (for example ref. 8, and reviewed in refs 9, 10), and the development of soluble *in vitro* systems capable of AAUAAA-dependent polyadenylation^{11–14} have allowed progress in the identification and characterization of factors required for authentic pre-mRNA polyadenylation (reviewed in ref. 15). Cleavage and polyadenylation of mRNA precursors require multiple protein factors including a poly(A) polymerase (PAP)^{16–20}. Initial experiments demonstrated that a nonspecific PAP activity could be separated from a cleavage-specificity factor, but that the PAP became specific for AAUAAA-containing RNAs when the fractions were combined¹⁶. Subsequently, it was shown that several highly purified PAPs that were nonspecific by themselves became AAUAAA-dependent when mixed with such a specificity factor^{21–24}. These findings

suggest that at least some of the previously identified PAPs can function in authentic pre-mRNA polyadenylation, and that PAP, like several other polymerases, acquires specificity through interactions with other proteins.

Here we describe the isolation of two types of cDNAs, apparently arising from alternative splicing, that can encode bovine PAP. *In vitro* translation of RNA transcribed from one of the cDNAs results in an 85K polypeptide that is active in nonspecific and AAUAAA-dependent polyadenylation. The deduced sequence of PAP reveals important insights into how the enzyme might function, and also suggests mechanisms by which PAP activity could be modulated.

Cloning of PAP cDNAs

With the aim of isolating a cDNA encoding PAP, we first demonstrated that a highly purified preparation of calf thymus PAP, isolated from a post-nuclear supernatant²⁵, was functional in AAUAAA-dependent polyadenylation when mixed with calf thymus specificity factor (results not shown). To determine whether PAP activity could be attributed to a single polypeptide, a sample of the PAP preparation was resolved by SDS gel electrophoresis. Staining of the gel with silver revealed one main protein band of ~55K, as well as several other bands of varying intensities (Fig. 1a). To determine if PAP activity could be attributed to a single polypeptide, conditions that allowed renaturation of nonspecific PAP activity were developed (see legend to Fig. 1). The results of such an analysis are shown in Fig. 1b, and indicate that only the 55K polypeptide had PAP activity. The sequences of three tryptic peptides from this protein were then determined (see legend to Fig. 2). Using oligonucleotide probes derived from one of these, 10 positive plaques were identified from a bovine heart library. All 10 cDNA inserts were

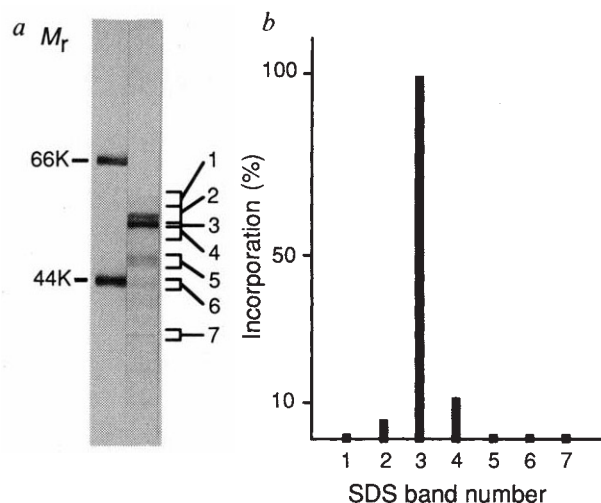


FIG. 1 PAP activity is contained in a single polypeptide of 55K. *a*, Highly purified poly(A) polymerase (10 μ g) was resolved on a 10% polyacrylamide SDS-gel and stained with silver. Visible bands (1–7) are indicated. *b*, Protein bands were eluted from a gel similar to that shown in *a*. Following renaturation, samples were tested in a nonspecific PAP assay. Results are shown as a percentage of the maximal incorporation (band 3), which was ~270,000 c.p.m., ~300-fold over background (940 c.p.m.).

METHODS. Calf thymus PAP (10 μ g), purified as described²⁵, was resolved on a 10% SDS gel, stained with copper and destained as in ref. 45. Elution of the bands and renaturation was done as described in ref. 46. Briefly, eluted proteins were precipitated with acetone, dissolved in 6 μ l 8M guanidine HCl and diluted with 100 μ l dilution buffer⁴⁶. After 2–5 h, 30 μ l was used in a nonspecific polyadenylation assay, done in 100 μ l reaction mixtures containing 0.5 MnCl₂, 0.3% polyvinyl alcohol (PVA), 0.1 mM ATP, 1 μ Ci α -[³²P]ATP, 100 mM Tris, pH 8.4, 0.1% BSA, 0.17 M guanidine-HCl, 50 mM NaCl, 7% glycerol and 300 ng 5' AAAAA-OH as primer. After incubation at 30 °C for 2 h, reaction mixtures were spotted on DEAE paper, dried briefly and washed twice with 0.2 M K₂HPO₄. Incorporation of [³²P]ATP was measured by scintillation counting.

related and several were similar in size. The nucleotide sequence of the longest was determined (Fig. 2a).

The PAP cDNA is 3,431 base pairs (bp) long, which for reasons detailed below is likely to be nearly full length. PAP cDNA contains one long open reading frame of 2,067 bp, capable of encoding a 689-residue polypeptide with a predicted M_r of 76,987. This M_r is considerably larger than predicted from the apparent size of the purified PAP. We believe this reflects limited proteolysis of the protein during purification and/or storage, but cannot exclude the possibility that the calf thymus PAP represents a different form of the enzyme, arising for example by alternative splicing. The cDNA sequence indicates that PAP mRNA contains a 160-nucleotide (nt) 5' untranslated leader that is extremely G-C rich (69% G+C). This feature, which has been observed in various housekeeping and growth-control genes (for example, see ref. 26), may restrict translation *in vivo* through the formation of stable secondary structure. By contrast, the 3' untranslated region (UTR), ~1,200 nt, is very A-U rich (67% A+U). A-U-rich sequences in 3' UTRs have been found in various lymphokine, cytokine and proto-oncogene mRNAs, and in several cases are responsible for destabilization of the mRNAs (for example refs 27–29). A consensus motif, UUAUUUAU, has been described in a number of lymphokine mRNAs³⁰, and may form a destabilizing signal sequence. PAP mRNA contains 19 matches (minimum six out of eight) to this consensus in its 3' UTR. An AAUAAA sequence is located 18 nts from the end of the cDNA. Although a poly(A) tract was not detected in this cDNA (or any of the others), the location of this sequence makes it likely to be the poly(A) signal.

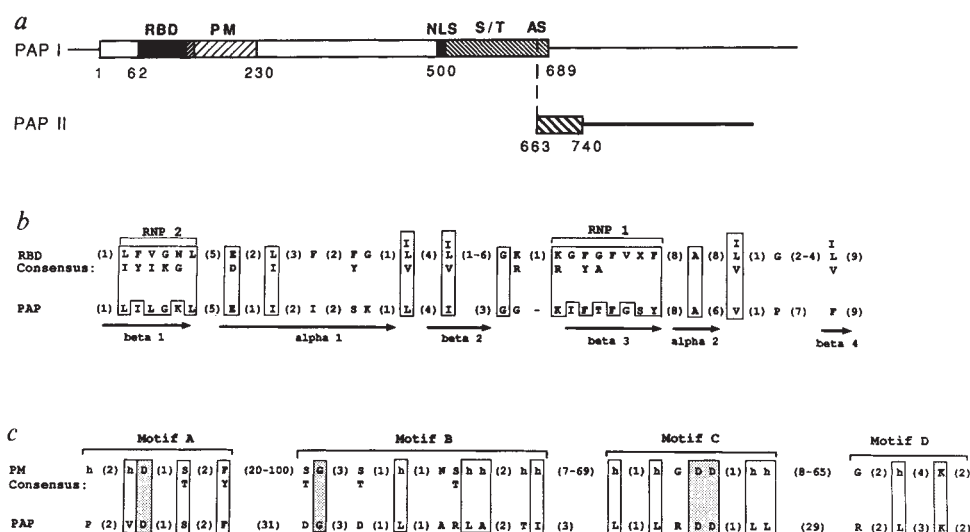
During analysis of the 10 cDNA clones, it became clear that they could be divided into two classes. Four, called class I PAP, are exemplified by the sequence shown in Fig. 2a. Six others (class II PAP) diverge near the 3' end of the coding region (at codon 663) and all encode a polypeptide of 740 residues or M_r 82, 838. The amino-acid sequences in the unique regions (Fig. 2b) do not reveal any features that suggest functional differences. But inspection of the roughly 900-nt 3' UTR of a class II cDNA does reveal an interesting feature: it is nearly devoid of A-U destabilizing elements. Despite the fact that the class II 3' UTR is 60% A+U, it contains only five matches (and none better than six out of eight) to the consensus mentioned above. This raises the novel possibility that alternative splicing might not significantly affect PAP itself, but rather produce mRNAs of differing stabilities.

Structural features of PAP

Comparison of the amino-acid sequence of class I (or class II) PAP with sequences in the GenBank database did not indicate any significant homologies. But, visual inspection of the sequence revealed several motifs that are likely to be of functional significance (Fig. 3a). First, sequences with considerable similarity to the RNP-consensus RNA-binding domain, (RBD; ref. 31) or RNA recognition motif³², can be found between residues 62 and 138 (Fig. 3b). This roughly 80 residue domain has been detected in a wide variety of proteins known to bind RNA, and where studied, binding is mediated by the RBD (reviewed in ref. 33). The most highly conserved and characteristic element of the RBD is the RNP-1 consensus octamer^{34,35}. A five out of eight match to this sequence is found between residues 96 and 103 in PAP, including a highly conserved basic residue at position 1 of the octamer and important aromatic residues at positions 3, 5 and 8. As the spacing between the RNP 1 and 2 motifs fits the consensus very well, and because a significant number of the dozen or so other highly conserved residues in the RBD are present in PAP, as are possible structural motifs³³, we conclude that PAP is a member of the RBD family of proteins.

Overlapping the C terminus of the RBD, extending roughly from residues 120 to 230, are sequences that bear strong similarity to the so-called RNA-dependent polymerase domain (reviewed in refs 36, 37). This putative domain, which extends

FIG. 3 Structural features of poly(A) polymerase. *a*, Block diagrams indicating structural features of PAP I and II. Class I and II PAP cDNAs contain identical 5' untranslated regions (UTR), but differ in their 3' UTRs, as described in the text. A possible RNA-binding domain (RBD), polymerase module (PM), and nuclear localization signal (NLS) as well as serine/threonine-rich regions (S/T) are indicated. The position of the putative alternative splice site (AS) is shown. Numbers, amino-acid residues. *b*, Comparison of the PAP RBD (residues 62–138) and the RBD consensus. Highly conserved residues, which were deduced from compilations presented in refs 31 and 32, are indicated. Positions at which identities between the consensus and PAP occur are boxed. The number of residues separating conserved positions are indicated, as are the RNP-1 (octamer) and RNP-2 consensus sequences. Possible positions of two α helices and four β strands, predicted from the known structure of one RBD (refs 49, 50) and the positions of certain conserved residues³³, as well as from Chou-Fasman analysis, are indicated by arrows. *c*, Comparison of the PAP polymerase module (PM; residues 121–230) and the PM consensus. Highly conserved residues, taken from compilations and consensus sequences



over about 100–200 residues, is found in many viral and cellular polymerases, including reverse transcriptases and the protein component of telomerase. It consists of four motifs separated by variable numbers of amino acids, and contains five essentially invariant residues and a number of other conserved amino acids (Fig. 3c). The PAP sequences indicated constitute a good match to this consensus, with the five invariant residues present and appropriate spacing separating the four motifs. Motif C has been suggested to have an important role in catalysis, and this consensus, or a variant of it, has been found in a variety of DNA-dependent as well as RNA-dependent polymerases³⁸. Together, these residues may constitute a polymerase module³⁶ that forms the enzyme's catalytic centre.

A putative nuclear localization signal (NLS) is found between residues 489 and 507. This sequence is an excellent match to the recently described bipartite NLS found in nucleoplasmin, which consists of two clusters of basic amino acids separated by about 10 spacer residues³⁹. Beginning precisely at the NLS and continuing to the C terminus of both classes of PAP are regions highly enriched in S and T residues, which constitute potential sites of post-translational modification. In the case of class I PAP, this region consists of 180 residues that are 20% S and 12% T. Particularly notable are three clustered sites (at residues 606, 648 and 654) that are potential targets for the cdc2 protein kinase^{40,41}, raising the possibility that PAP activity may be subject to cell-cycle control.

Expression of PAP *in vitro*

To analyse the structure and function of PAP, we developed conditions that allow synthesis of active PAP by translation in a reticulocyte lysate (see legend to Fig. 4). Figure 4a displays the [³H]leucine-labelled proteins synthesized from a template encoding full-length class I PAP (lane 2), as well as from several 3' truncated templates (lanes 3–5). The apparent size of PAP estimated from its gel mobility is about 85K, somewhat larger than its predicted mass of 77K.

To determine whether the full-length polypeptide or any of the truncated derivatives produced by *in vitro* translation possess PAP activity, we first tested the translation products in a non-specific assay essentially identical to the one used in the experiment shown in Fig. 1. The results (Fig. 4b) indicate that full-length (689 amino acids) PAP was quite active (lane 2), giving

derived from RNA-dependent RNA polymerases in refs 36 and 37, as well as from rules for motif C described in ref. 38, are indicated. Positions at which identities occur are boxed, and the five invariant residues are shaded. The number of residues separating conserved positions are indicated. Any hydrophobic amino acid.

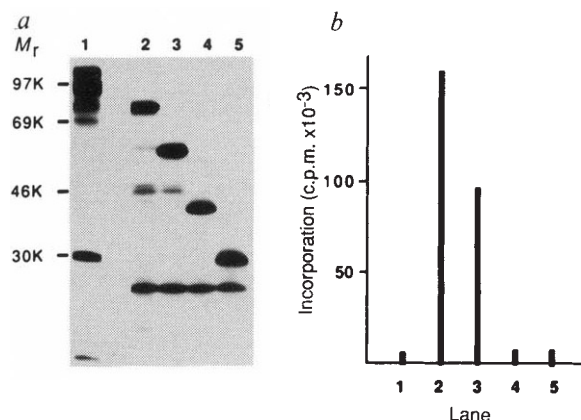


FIG. 4 *In vitro* expression of PAP. *a*, *In vitro* translation. PAP RNA was produced by *in vitro* transcription with SP6 RNA polymerase from a full-length PAP cDNA template or templates cleaved with *SpeI*, *ApaLI* or *BaniI*. RNAs were translated in a reticulocyte lysate in the presence of [³H]leucine, proteins were resolved on a 12% polyacrylamide-SDS gel, and products visualized by fluorography. RNAs used, and the predicted length of the protein product, were: Lane 1, Brome mosaic virus RNA; lane 2, full-length, 689 amino acids; lane 3, *SpeI*, 538 amino acids; lane 4, *ApaLI*, 379 amino acids; and lane 5, *BaniI*, 308 amino acids. The positions of protein standards are shown on the left. *b*, Nonspecific polyadenylation assay with *in vitro* translation products. Aliquots (5 µl) of the *in vitro* translation reactions shown in *a* were added to 100 µl reaction mixtures, processed and quantitated as described in the legend to Fig. 1. Numbers on the horizontal axis correspond to the lane numbers in *a*. Incorporation by full-length PAP (160,000 c.p.m.) was ~30-fold over background (5,400 c.p.m.). Bars are averages of two independent experiments.

METHODS. Templates for *in vitro* transcription were prepared by PCR, using sense and antisense primers complementary to the 5' and 3' ends of the coding region and the PAP4 cDNA as template, as described by R. Aurora and W. Herr (manuscript submitted). The PCR product generated contained the SP6 promoter, nucleotides from –11 to +4 (relative to the AUG) of the rabbit β -globin gene, and the entire PAP open reading frame. Amino acid 2 was changed from Pro to Ala. The resulting linear DNA was cleaved with the restriction enzymes indicated and used in *in vitro* transcription reactions as described⁵¹. *In vitro* translation was done as suggested by the suppliers of the lysate (Promega).

rise to a roughly 30-fold increase over the background incorporation obtained with control translations that had contained brome mosaic virus (BMV) RNA (lane 1). Substrate specificity was also observed with *in vitro* translated PAP, as [32 P]dATP, UTP or GTP gave no incorporation above background, nor did reactions lacking a primer (results not shown). The 538 amino-acid PAP retained about 60% activity (lane 3), indicating that the S/T-rich region is not essential in this assay. By contrast, the 379 and 308 amino-acid derivatives (lanes 4 and 5) were inactive, indicating that sequences towards the C-terminal of the putative polymerase domain are in some way necessary for activity.

To determine whether the *in vitro* expressed PAP products analysed above were able to carry out AAUAAA-dependent polyadenylation, translation products were incubated with highly purified calf thymus specificity factor (SF), either wild-type or mutant 32 P-labelled pre-RNA, and RNA products were resolved by gel electrophoresis. The results (Fig. 5) indicate that the full-length 689-residue protein was indeed active in AAUAAA-dependent polyadenylation (lane 5). As with authentic PAP, activity was not detected in the absence of SF (lane 4) or with the AAAAAA-containing pre-RNA (lanes 6–8). Note that there was a low level of endogenous PAP activity in the reticulocyte lysate, which was also activated by SF (lane 3). This is consistent with the results of the nonspecific assay (Fig. 4b), which showed a low background incorporation. But full-length *in vitro* PAP resulted in substantially elevated activity, seen in both the increased length of the poly(A) tails and the greater fraction of pre-RNA polyadenylated (compare lanes 3 and 5). Interestingly, the 538-residue mutant derivative again retained substantial activity (lanes 9 and 10), whereas the 379- and 308-residue proteins were inactive (lanes 11–14). Together, these results show that PAP produced *in vitro* from a cloned cDNA is active in both nonspecific poly(A) synthesis and AAUAAA-dependent polyadenylation, and that the S/T-rich C-terminal region of the enzyme, although not essential, enhances both activities.

Discussion

We have described the isolation of two classes of cDNAs, apparently arising by alternative splicing, that encode highly related PAPs. The calculated sizes of the two polymerases (about 77 and 83K) are larger than previous estimates of purified PAPs (refs 2, 3). Although it is possible that some of the previously described forms may have arisen by different alternative splicing than detected here, or even by the product(s) of a distinct gene(s), we believe that the size discrepancies result from partial proteolysis during purification. The fact that about

20K could be deleted from the C terminus of cloned PAP without greatly decreasing activity is consistent with this possibility.

On the basis of the structural features of PAP, we can propose a rough model for how catalysis might occur. We suggest that the substrate RNA is recognized and bound by the N-terminal RBD. As purified PAP does not seem to display primer specificity, it may be that there is little, if any, nucleotide sequence-specificity in this interaction although it is perhaps likely that an RNA 3' end is an important element for recognition. We further suggest that the 3'OH of the primer is situated so that it is accessible to the closely juxtaposed polymerase module (PM), which we propose to be the actual site of catalysis. Sequences extending at least 150 residues from the C-terminal of the PM are also required for activity *in vitro*. Whether these residues have a direct role in polymerization or function in determining the overall structure of the protein remains to be determined. Finally, the S/T-rich region is not essential for function *in vitro*. We suggest that this region is regulatory in nature, and that its full function may involve interactions with components not present in the *in vitro* reactions.

The presence of putative RNA binding and polymerase domains in one protein has not, to our knowledge, been observed previously. Thus the fact that these clearly independent motifs seem to overlap partially in PAP is especially intriguing, and how such a structure may have evolved is an interesting question. The presence of a putative PM in PAP is itself significant and provides some insight into the function of this motif in other polymerases. All other polymerases containing the motif (more than 80 have been identified³⁶) are template-dependent. As PAP does not use a template, this argues that template binding is not a requisite function of the PM. The PAP polymerase domain is not a perfect match to the consensus (for example, a well conserved Gly immediately to the N-terminal side of the invariant Asp-Asp sequence in motif C is Arg in PAP, and the number of residues separating motifs B and C is low). This may reflect the lack of necessity of accommodating a template and/or the ability of PAP to use only ATP as a substrate.

It is perhaps not surprising that PAP activity should be subject to control, for example during cell growth or differentiation (see refs. 42, 43 for related reviews). But the number of possible post-transcriptional regulatory mechanisms suggested by the predicted mRNA and protein sequences is intriguing. In the mRNA, the G-C-rich 5' UTR may restrict translation through secondary structure formation, a phenomenon that can be subject to modulation under growth conditions⁴⁴. The large number of putative A-U-rich destabilizing elements in the class I 3' UTR may also modulate PAP production by facilitating mRNA

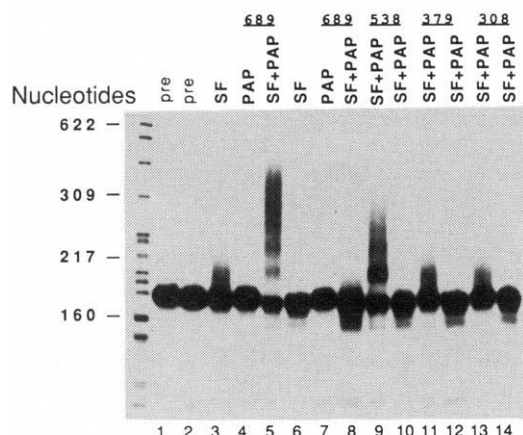


FIG. 5 PAP expressed *in vitro* can function in AAUAAA-dependent polyadenylation. Aliquots (5 μ l) of the *in vitro* translation reactions shown in Fig. 4a were added to 50 μ l reaction mixtures containing wild-type or mutant pre-mRNA and specificity factor as indicated, and the products were analysed by gel electrophoresis. Lane 1, AAUAAA-containing SV40 late RNA precursor (wt-pre); lane 2, AAAAAA-containing SV40 RNA (pm-pre). Samples analysed in lanes 3–5, 9, 11 and 13 contained wt RNA; 6–8, 10, 12 and 14 contained pm RNA. Translation products added to processing reactions were from translations that had contained the following mRNAs: lanes 3 and 4, BMV RNA; lanes 5–8, full-length PAP RNA; lanes 9 and 10, RNA from *SpeI*-cleaved template; lanes 11 and 12, RNA from *ApaI*-cleaved template; and lanes 13 and 14, RNA from *BanI*-cleaved template. Numbers at the top indicate the sizes of the resultant PAPs, in amino acids. Samples analysed in lanes 3, 5, 6 and 8–14 contained purified calf-thymus specificity factor (SF), and those in lanes 4 and 7 lacked SF. DNA markers are shown on the left. Sizes are in nucleotides.

METHODS. Calf thymus SF was purified to ~50% homogeneity as described (ref. 18 and K. Murthy and J.L.M., manuscript in preparation). SF (0.2 μ g) was added to reaction mixtures as indicated. Processing reactions were done using [32 P]GTP-labelled RNAs exactly as described previously²², an important point being that reaction mixtures contained $MgCl_2$, instead of $MnCl_2$, which prevents nonspecific polyadenylation. Reaction products were resolved on a 6% polyacrylamide-8M urea gel.

turnover and, as discussed above, different forms of the mRNA may have different stabilities. In the protein itself, the S/T-rich C-terminal region may have a regulatory role, as a target for

phosphorylation and/or other modifications. Future studies of these and other issues raised by the mRNA and protein sequences described here should be informative. □

Received 3 July; accepted 22 August 1991.

- Edmonds, M. & Abrams, R. J. *biol. Chem.* **235**, 1142–1149 (1960).
- Edmonds, M. in *The Enzymes* Vol 15 (ed. Boyer, P. D.) 217–244 (Academic, New York, 1982).
- Jacob, S. T. & Rose, K. M. in *Enzymes of Nucleic Acid Synthesis and Modification* Vol 2, (ed. Jacob, S. T.) 135–157 (CRC Press, Boca Raton, Florida, 1983).
- Darnell, J. E., Wall, R. & Tushinski, R. J. *Proc. natn. Acad. Sci. U.S.A.* **68**, 1321–1325 (1971).
- Lee, S., Mendecki, J. & Brawerman, G. *Proc. natn. Acad. Sci. U.S.A.* **68**, 1331–1335 (1971).
- Edmonds, M., Vaughan, M. & Nakazato, H. *Proc. natn. Acad. Sci. U.S.A.* **68**, 1336–1340 (1971).
- Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211–214 (1976).
- Fitzgerald, M. & Shenk, T. *Cell* **24**, 251–260 (1981).
- Manley, J. L. *Biochem. biophys. Acta* **950**, 1–12 (1988).
- Humphrey, T. & Proudfoot, N. J. *Trends Genet.* **4**, 243–245 (1988).
- Manley, J. L. *Cell* **33**, 595–605 (1983).
- Manley, J. L., Yu, H. & Ryner, L. C. *Molec. cell. Biol.* **5**, 373–379 (1985).
- Moore, C. L. & Sharp, P. A. *Cell* **36**, 581–591 (1984).
- Moore, C. L. & Sharp, P. A. *Cell* **41**, 845–855 (1985).
- Wickens, M. *Trends biochem. Sci.* **15**, 277–281 (1990).
- Takagaki, Y., Ryner, L. C. & Manley, J. L. *Cell* **52**, 731–742 (1988).
- Christofori, G. & Keller, W. *Cell* **54**, 875–889 (1988).
- Takagaki, Y., Ryner, L. C. & Manley, J. L. *Genes Dev.* **3**, 1711–1724 (1989).
- Gilmartin, G. M. & Nevins, J. R. *Genes Dev.* **3**, 2180–2189 (1989).
- Takagaki, Y., Manley, J. L., MacDonald, C. C., Wilusz, J. & Shenk, T. *Genes Dev.* **4**, 2112–2120 (1990).
- Christofori, G. & Keller, W. *Molec. cell. Biol.* **9**, 193–203 (1989).
- Ryner, L. C., Takagaki, Y. & Manley, J. L. *Molec. cell. Biol.* **9**, 4229–4238 (1989).
- Bardwell, V. J., Zarkower, D., Edmonds, M. & Wickens, M. *Molec. cell. Biol.* **10**, 846–849 (1990).
- Wahle, E. *J. biol. Chem.* **266**, 3131–3139 (1991).
- Tsiapalis, C., Dorson, J. & Bollum, F. J. *J. biol. Chem.* **250**, 4486–4496 (1975).
- Kozak, M. *J. Cell Biol.* **108**, 229–241 (1989).
- Shaw, G. & Kamen, R. *Cell* **46**, 659–667 (1986).
- Jones, T. R. & Cole, M. D. *Molec. cell. Biol.* **7**, 4513–4521 (1987).

- Wilson, T. & Treisman, R. *Nature* **336**, 396–399 (1988).
- Caput, D. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1670–1674 (1986).
- Bandziulis, R. J., Swanson, M. S. & Dreyfuss, G. *Genes Dev.* **3**, 431–437 (1989).
- Query, C. C., Bentley, R. C. & Keene, J. D. *Cell* **57**, 89–101 (1989).
- Kenan, D. J., Query, C. C. & Keene, J. D. *Trends biochem. Sci.* **16**, 214–220 (1991).
- Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K. & Dreyfuss, G. *Molec. cell. Biol.* **6**, 2932–2943 (1986).
- Sachs, A. B., Bond, M. W. & Kornberg, R. D. *Cell* **45**, 827–835 (1986).
- Poch, O., Sauvaget, I., Delaure, M. & Tordo, N. *EMBO J.* **8**, 3867–3874 (1989).
- Lundblad, V. & Blackburn, E. H. *Cell* **60**, 529–530 (1990).
- Argos, P. *Nucleic Acids Res.* **16**, 9909–9916 (1988).
- Robbins, J., Dilworth, S. M., Laskey, R. A. & Dingwall, C. *Cell* **64**, 615–624 (1991).
- Langan, T. A. *J. biol. Chem.* **257**, 14835–14846 (1982).
- Shenoy, S. et al. *Cell* **57**, 763–774 (1989).
- Jackson, R. J. & Standart, N. *Cell* **62**, 15–24 (1990).
- Wickens, M. *Trends biochem. Sci.* **15**, 320–324 (1990).
- Kozak, M. *Molec. cell. Biol.* **8**, 2737–2744 (1988).
- Lee, C. et al. *Analyt. Biochem.* **166**, 308–312 (1987).
- Hager, D. & Burgess, R. *Analyt. Biochem.* **109**, 76–86 (1980).
- Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. H. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6970–6974 (1987).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. *Nature* **348**, 515–520 (1990).
- Hoffman, D. W., Query, C. C., Golden, B., White, S. W. & Keene, J. D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2495–2499 (1991).
- Melton, D. A. et al. *Nucleic Acids Res.* **12**, 7035–7056 (1984).

ACKNOWLEDGEMENTS. We thank L. Zhong for technical assistance, R. Aurora for describing the method used to prepare templates for *in vitro* translation, K. Murthy for highly purified specificity factor and K. Murthy, H. Ge, J. Harper and Y. Takagaki for discussions. This work was supported by the NIH. T.R. was supported by a fellowship from the DAAD during the early phases of this work.

LETTERS TO NATURE

Is the Galactic Centre gamma-ray source 1E1740.7–2942 accreting from a molecular cloud?

John Bally* & Marvin Leventhal†

* AT&T Bell Laboratories, HOH L-245, Holmdel, New Jersey 07733, USA

† AT&T Bell Laboratories, MH 1E-349, 600 Mountain Avenue, Murray Hill, New Jersey, USA

FOR twenty years there have been detections of an intermittent source of 511-keV electron–positron annihilation radiation at the centre of our Galaxy^{1–3}. Until recently, none of the balloon or satellite experiments lucky enough to catch the source in its ‘on’ state had spatial resolution better than several degrees, but on 13–14 October 1990 the French SIGMA experiment aboard the Soviet GRANAT spacecraft observed a day-long burst of narrow-band γ rays, probably annihilation radiation, from within 1.5 arcmin of the known X-ray source 1E1740.7–2942, about 50 arcmin from Sgr A West, which contains the dynamical centre of the Milky Way^{4–6}. Here we report millimetre-wavelength observations showing that this variable X-ray source lies on a line of sight to the dense $10^5 M_\odot$ molecular cloud G–0.86–0.08, near the Galactic Centre. We suggest that the γ -ray source, probably a compact object, possibly a black hole, is accreting dense gas directly from the molecular cloud.

SIGMA observed the Galactic Centre region for ten days in 1990. For one day a strong transient feature peaking near 511 keV with a high-energy cutoff above 511 keV and an extended low-energy tail was detected, but had disappeared by the next day. Given the statistical uncertainty and the 40-keV resolution of SIGMA, the data are consistent with positron-annihilation line radiation plus a low-energy tail produced by Compton scattering, or alternatively with annihilation from positronium (refs 4, 5, E. P. Liang and C. D. Dermer, preprint), a bound state of e^+e^- . With 40-keV resolution, the singlet positronium annihilation line is expected to be only moderately brighter than

the adjacent triplet positronium low-energy continuum⁷. The intensity of this feature is consistent with earlier observations of the Galactic Centre positron source. It is therefore tempting to identify 1E1740.7–2942 as the Galactic Centre positron source. But because a sharp 511-keV line has not been unambiguously detected, some caution is warranted. Other transient 511-keV line-radiation sources have been recently identified in other parts of the sky including Nova Muscae 1991⁸ and V1223 Sagittarii⁹. The possibility that several independent objects contribute to the Galactic Centre positron source cannot therefore be ruled out.

The position of the γ -ray source was determined by SIGMA’s coded aperture system which located the source to within an error circle with a radius of 1.5 arcmin. Many workers believe that the positrons are produced by an accreting black-hole system, which should also be an X-ray source. The γ -ray source may therefore be coincident with the hard X-ray source 1E1740.7–2942 whose position is known to within 22 arcsec. 1E1740.7–2942 was first detected by the Einstein Observatory as one of a large group of weak X-ray sources surrounding Sgr A West^{10,11}. Subsequent hard X-ray^{12,13} and γ -ray¹⁴ experiments have established that at times, this is the dominant source of hard radiation within several degrees of the Galactic Centre.

A number of factors have led theorists to believe that a black hole of modest mass ($1\text{--}1,000 M_\odot$) is involved in the production of positrons^{7,15}. In the ‘on’ state, $\sim 10^4 L_\odot$ of annihilation radiation is produced in a region $< 10^{15}$ cm in extent, because source exhibits day-to-day variability. At other times, a ‘Comptonized disk’ spectrum of hard X-rays^{16,17} very similar to that produced by the stellar-size black-hole candidate Cyg X-1 is observed. It is believed that the positrons are produced by γ - γ collisions in a relativistic plasma which is occasionally formed around the black hole. The positrons must escape the plasma, slow down and then annihilate in a cold ($T < 10^5$ K) stopping medium in order to produce the sharp 511-keV line previously reported^{2,18–20}. If the positron energies are in the keV to MeV range, as expected from most production mechanisms, then particle densities $> 10^5 \text{ cm}^{-3}$ are required to moderate and stop the positrons and have them annihilate on a < 1 -year timescale¹⁹;

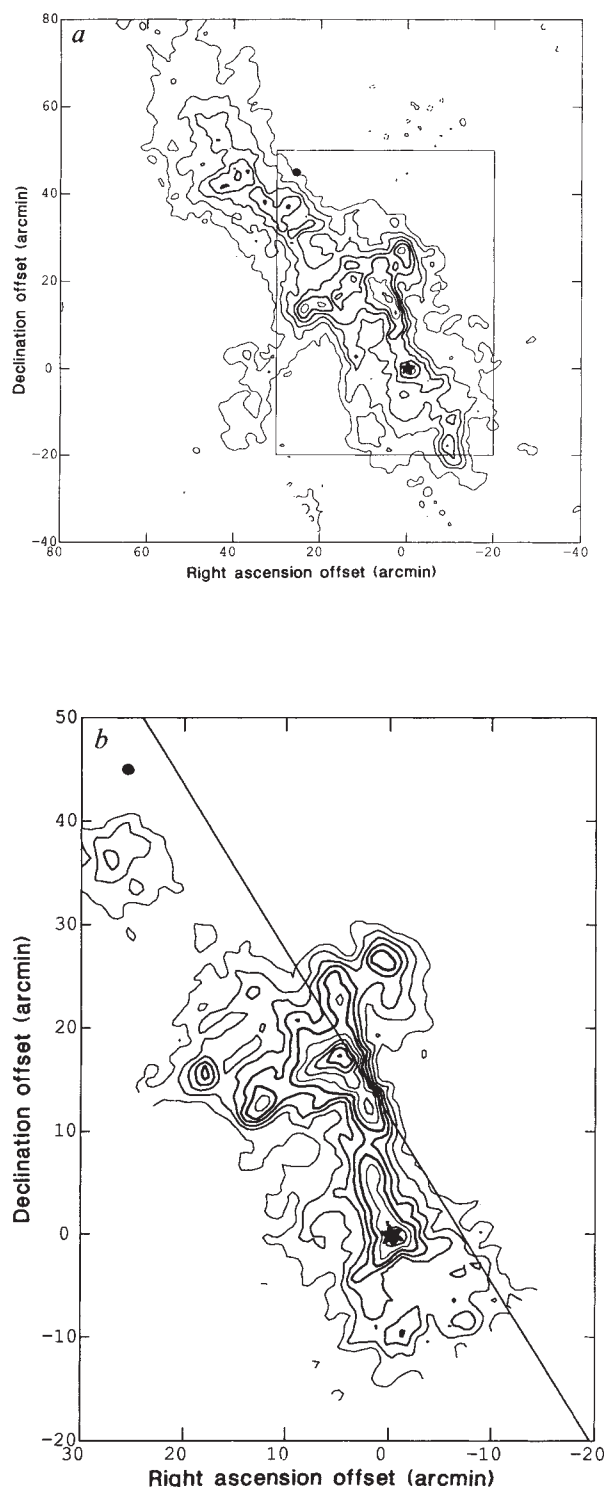


FIG. 1 *a*, Contour map of the ^{12}CO emission from the Galactic Centre region showing the gas in the velocity range $V_{\text{LSR}} = -200$ to -100 km s^{-1} with slightly degraded (2 arcmin) resolution. The centre of the map at (0,0) corresponds to $\alpha(1950) = 17 \text{ h } 40 \text{ min } 41.4 \text{ s}$, $\delta(1950) = -29^\circ 43' 21.0''$, the position of the hard X-ray source 1E1740.7-2942 (marked with a * symbol). The position of the Galactic Centre is shown by a filled circle near the upper left corner of the box. The box shows the spatial extent of the ^{13}CO map shown in (*b*). Contour levels are displayed at intervals of 75 K km s^{-1} . *b*, A contour map of ^{13}CO emission around the position of 1E1740.7-2942. The integration velocity range extends from $V_{\text{LSR}} = -200$ to -100 km s^{-1} . Contours are displayed every 10 K km s^{-1} , starting at 20 and extending to 110 K km s^{-1} . Symbols and lines are the same as in (*a*). The diagonal line is the Galactic equator ($b=0$).

the minimum density increases inversely with the variability timescale. A stellar wind from a massive companion star might provide a natural stopping medium. Although there is as yet no evidence for such a stellar companion, and the large extinction towards the source will hamper the search, a binary model for the γ -ray source cannot be ruled out.

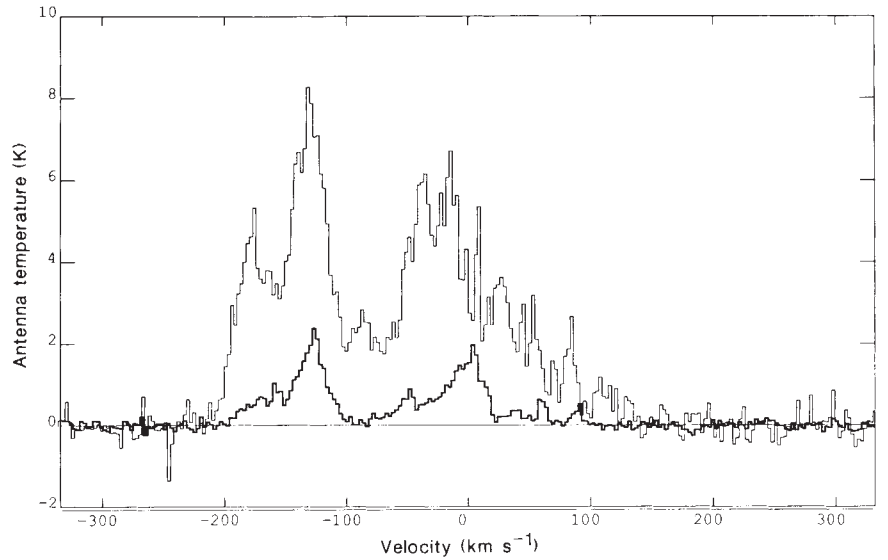
Here we report new millimetre-wave observations which indicate that 1E1740.7-2942 is positioned within 1 arcmin of the centre of a Galactic Centre molecular cloud. Instead of being fed by a stellar companion, the compact object may be accreting gas directly from the molecular cloud. The Crawford Hill 7-m millimetre-wave antenna was used to map the distribution of molecular gas over the entire Galactic Centre region. The large-scale distribution of the 98-GHz transition of CS, the 110-GHz transition of ^{13}CO , and the 115-GHz transition of ^{12}CO has been discussed by Bally *et al.*^{21,22}. Recently, the maps were extended to include negative longitudes down to $l = -2.2^\circ$. The antenna has a beam size of 100 arcsec and all data presented here were obtained on a uniform 1-arcmin grid in galactic coordinates (l, b). All observations were obtained with a SIS (superconductor-insulator-superconductor junction) receiver operating in single-sideband mode, with a receiver temperature T_{SSB} of 65–150 K. The data were obtained with a 512-channel, 1-MHz-per-channel filterbank spectrometer, giving a velocity resolution of $\sim 2.6 \text{ km s}^{-1}$. Details of the Galactic Centre mapping and the large-scale characteristics of the molecular gas distribution in this part of the sky are discussed by Bally *et al.*^{21,22}.

Figure 1*a* and *b* shows the spatial distribution of molecular gas in the vicinity of 1E1740.7-2942 in both ^{12}CO and ^{13}CO . As shown in the spectra presented in Fig. 2, two major cloud velocity components are seen towards the γ -ray source. The velocity component centred at $V_{\text{LSR}} = -130 \text{ km s}^{-1}$ (where LSR stands for local standard of rest) exhibits a highly localized maximum in the integrated intensity in both ^{12}CO and ^{13}CO at the position of 1E1740.7-2942 and extends over a 140 km s^{-1} range from $V_{\text{LSR}} = -200$ to -60 km s^{-1} . This CO cloud is the brightest molecular feature within an 8-arcmin region at any velocity along this line of sight. This molecular peak lies in a relatively clear portion of the Galactic Centre at the extreme low-longitude end of a ridge of molecular emission which can be traced towards Sgr C and Sgr A. Outside this velocity range ($V_{\text{LSR}} = -200$ to -60 km s^{-1}), strong emission ($> 2 \text{ K}$ in ^{12}CO) is found only between $V_{\text{LSR}} = -50$ and 30 km s^{-1} .

There is some probability of a chance superposition of a random cloud and the γ -ray source. We can estimate the likelihood from the area-filling factor of bright CO emission in this part of the sky. Although there is faint emission throughout the region, we estimate that at the intensity level found within 1 arcmin of the γ -ray source ($> 400 \text{ K km s}^{-1}$), $< 5\%$ of the sky is covered at any velocity within a scale height of the galactic equator. A crude estimate of the probability of a chance association is $\sim 5\%$.

The gas centred at -130 km s^{-1} is in the negative longitude portion of the Galactic Centre molecular disk^{21,22} which contains the densest and most massive cloud complexes near the Galactic Centre. The large negative velocity at this longitude suggests that its true galactocentric distance is comparable to its projected distance, $\sim 120 \text{ pc}$ from the Galactic Centre (assumed to lie at a distance of $\sim 7.5 \text{ kpc}$). Tidal shear in the nearly flat rotation curve of the Galaxy constrains the mean density of self-gravitating clouds to have an average volume density $n(\text{H}_2) > 2 \times 10^4 [R/100 \text{ pc}]^{-2} \text{ cm}^{-3}$ (R is the distance from the centre of the Galaxy). We have detected highly localized CS $J=2-1$ emission from this cloud with an integrated intensity of 13 K km s^{-1} between $V_{\text{LSR}} = -200$ and -100 km s^{-1} , which requires gas densities of this order to be excited into emission. This cloud has properties similar to other molecular clouds in the Galactic Centre; the line width ($> 50 \text{ km s}^{-1}$) and mean density ($> 10^4 \text{ cm}^{-3}$) are an order of magnitude larger than is typical for clouds near the Sun.

FIG. 2 The ^{12}CO (light) and ^{13}CO (heavy) line profiles averaged over a 1.5-arcmin field of view centred on the position of 1E1740.7–2942. The component centred at $V_{\text{LSR}} = -130 \text{ km s}^{-1}$ is the feature that peaks at the position of the positron source. The other velocity components have extended spatial structure and do not exhibit spatial coincidence with the γ -ray source.



The other cloud components shown in Fig. 2 trace molecular gas which lies mostly outside the Galactic Centre region. None of these other velocity components shows any obvious spatial relationship with 1E1740.7–2942. The broad feature centred near $V_{\text{LSR}} = 0 \text{ km s}^{-1}$ consists of a blend of foreground and background gas in the disk of the Galaxy and is dominated by emission lying within a few kpc of the Sun. The spike near $V_{\text{LSR}} = -60 \text{ km s}^{-1}$ is associated with the 3-kpc spiral arm in the disk. The ^{12}CO line profile shows components at $V_{\text{LSR}} = -200$ to -160 km s^{-1} , and also at 20 to 120 km s^{-1} . This emission is produced by the so-called expanding molecular ring (EMR) which consists of diffuse, lower-density ($n(\text{H}_2) \approx 10^3 \text{ cm}^{-3}$) molecular gas and which may be located between 200 and 500 pc from the Galactic Centre.

The integrated line area of the ^{13}CO spectrum between $V_{\text{LSR}} = -200$ and -100 km s^{-1} is 82 K km s^{-1} , which for an assumed excitation temperature of 15 K implies an LTE (local thermodynamic equilibrium) column density of $N(^{13}\text{CO}) \approx 6 \times 10^{16} \text{ cm}^{-2}$. Using the empirically determined ratio, $N(\text{H}_2)/N(^{13}\text{CO}) = 7 \times 10^5$ (calibrated in the local interstellar medium) gives $N(\text{H}_2) \approx 4 \times 10^{22} \text{ cm}^{-2}$. We can also estimate the molecular-cloud column density from the standard conversion^{23,24} of $N(\text{H}_2)/\int T_A^* dV = 2.6 \times 10^{20} \text{ cm}^{-2}/\text{K km s}^{-1}$ (where T_A^* is antenna brightness temperature corrected for atmospheric attenuation), derived for clouds near the Sun by comparison of the 100-MeV γ -ray flux with the average ^{12}CO surface brightness. The observed ^{12}CO flux integrated from $V_{\text{LSR}} = -200$ to -60 km s^{-1} ($=530 \text{ K km s}^{-1}$) implies $N(\text{H}_2) = 1.4 \times 10^{23} \text{ cm}^{-2}$ for the cloud associated with 1E1740.7–2942.

The total column density along the line of sight, derived by integrating over all ^{12}CO emission shown in Fig. 2, is estimated to be $N(\text{H}_2) = 3 \times 10^{23} \text{ cm}^{-2}$, corresponding to a visual extinction of $A_V \approx 200$ magnitudes. The column density required to cut off the low-energy end of the X-ray spectrum¹² from 1E1740.7–2942 is $N(\text{H}) \approx 1.4 \times 10^{23} \text{ cm}^{-2}$ (or $N(\text{H}_2) \approx 7 \times 10^{22} \text{ cm}^{-2}$), in reasonable agreement with our estimate of the column density. Metallicity effects and variations in the physical environment along the line of sight make these estimates uncertain by at least a factor of two. The large column density towards 1E1740.7–2942 makes detection at optical or even near-infrared wavelengths unlikely and may explain the low energy cutoff in the X-ray spectrum.

The total mass of gas associated with the cloud between $V_{\text{LSR}} = -160$ to -100 km s^{-1} within a 4-arcmin (9-pc) radius is estimated to be $\sim 10^5 M_\odot$ (using the LTE method discussed above). When corrected for the 100 arcsec beam size, the cloud has a mean radius of $\sim 4.5 \text{ pc}$, and the mean volume averaged

density is estimated to be at least $n(\text{H}_2) \approx 5 \times 10^3 \text{ cm}^{-3}$ near that required to resist tidal disruption. Most molecular clouds have a volume-filling factor of less than 0.1, so the average gas density is likely to be at least $5 \times 10^4 \text{ cm}^{-3}$, typical of clouds in the Galactic Centre disk population and sufficient to explain the CS emission.

We now estimate the accretion luminosity onto a compact object moving through a dense molecular cloud. We use the formulation first discussed by Bondi and Hoyle^{25,26} and assume that the accretion is supersonic. All matter flowing past the source with velocity v and an impact parameter less than the accretion radius r_a , given by the condition that $v^2 = GM_*/r_a$ (where M_* is the mass of the accreting object), falls onto the compact object. The gravitational field of the object deflects the passing matter towards the axis of motion. A shock forms behind the compact object where streamlines cross, dissipating some of the gravitational potential energy. An accretion column forms, which in case of a perfectly steady flow can directly fall onto the compact object. The structure of the accretion flow near the compact object should be similar to the flows from companion stars, although in our case the angular momentum of the accreted material is likely to be much smaller than in a binary star. A fraction η of the rest mass energy released by accreting matter is converted into the source luminosity. The accretion luminosity is therefore given by $L = \eta \dot{M} c^2$ where $\dot{M} = \pi \rho r_a^2 v$ (where ρ is the mass density of the cloud far from the gravitational influence of the accreting object). Thus,

$$L = \eta \pi \mu m(\text{H}_2) n(\text{H}_2) G^2 M_*^2 c^2 / v^3 \\ = 2.3 \times 10^{36} \eta \left[\frac{M}{M_\odot} \right]^2 \left[\frac{v}{10 \text{ km s}^{-1}} \right]^{-3} \\ \times \left[\frac{n}{10^4 \text{ cm}^{-3}} \right] \quad (\text{erg s}^{-1})$$

where $\mu = 1/1 - Y = 1.4$ ($Y = 0.28$ is the mass fraction of helium in interstellar gas). The observed γ -ray luminosity ranges from a high state value of $\sim 10^{37} \text{ erg s}^{-1}$ to a low state value of $\sim 10^{35} \text{ erg s}^{-1}$. If the average luminosity is $10^{36} \text{ erg s}^{-1}$, and this represents 10% of the total source luminosity, then a $2M_\odot$ compact object moving at a relative velocity of 10 km s^{-1} with respect to the cloud can provide the luminosity. A random star moving with respect to a random cloud near the Galactic Centre is likely to have a relative velocity of ~ 10 – 100 km s^{-1} . For a relative velocity of 100 km s^{-1} , a $60M_\odot$ object would produce the same luminosity.

The accretion flow is likely to be unsteady, providing a natural

explanation for the time-variability of the γ -ray source. For a neutron star or black hole, the ratio of the accretion radius to the Schwarzschild radius is in the range 10^7 – 10^9 , so tiny fluctuations in the velocity field of the cloud can provide enough residual angular momentum to produce instabilities in the accretion flow near the compact object, resulting in unsteady accretion. Furthermore, for an isothermal fluid, shocks compress the gas flowing through the accretion column²⁷. Therefore, dissipation in the accretion flow is likely to lead to a gas density $n(\text{H}_2) > 10^7 \text{ cm}^{-3}$ near the compact object, sufficiently to moderate the positron velocities and act as a stopping medium.

We propose that the Galactic Centre γ -ray source 1E1740.7–2942 may be produced by the chance encounter of a massive stellar-remnant black hole with a dense Galactic Centre molecular cloud. Such sources are most likely to be visible near the Galactic Centre owing to the increased volume density of molecular gas in this region ($\sim 10\%$ of the total mass is in the interstellar medium in the inner 500 pc as opposed to $< 1\%$ in the vicinity of the Sun) which raises the probability of a significant accretion inducing encounter. Furthermore, the higher mean density of clouds in the inner Galaxy increases the expected accretion luminosity. $\square$

Received 25 March; accepted 26 July 1991.

1. Lingenfelter, R. E. & Ramaty, R. *Astrophys. J.* **343**, 686–695 (1989).
2. Leventhal, M. *et al. Nature* **339**, 36–38 (1989).
3. McClintock, J. E. & Leventhal, M. *Astrophys. J.* **346**, 143–150 (1989).
4. Paul, J. *et al. in Proc. Int. Symp. on Gamma-Ray-Line Astrophysics, Saclay 1990* (eds Prantzos, N. & Durouchoux, P.) (American Institute of Physics, New York, in the press).
5. Sunyaev, R. *et al. in Proc. Int. Symp. on Gamma-Ray-Line Astrophysics, Saclay 1990* (eds Prantzos, N. & Durouchoux, P.) (American Institute of Physics, New York, in the press).
6. Sunyaev, R. *et al. Soviet Astr. Lett.* **17**, 126–129.
7. Leventhal, M. *Astrophys. J.* **183**, L147–L150 (1973).
8. Sunyaev, R., Jourdain, E. & Goldwurm, A. *IAU Circ.* No 5201 (1991).
9. Briggs, M., Matteson, J., Gruber, D. & Peterson, L. *IAU Circ.* No 5229 (1991).
10. Watson, M. G., Willingale, R., Grindlay, J. E. & Hertz, P. *Astrophys. J.* **250**, 142–154 (1981).
11. Hertz, P. & Grindlay, J. E. *Astrophys. J.* **278**, 137–149 (1984).
12. Skinner, G. K. *et al. Nature* **330**, 544–547 (1987).
13. Kawai, N. *et al. Astrophys. J.* **330**, 130–141 (1988).
14. Cook, W. R. *et al. Astrophys. J.* **372**, L75–L78 (1991).

15. Lingenfelter, R. E. & Ramaty, R. in *Positron–Electron Pairs in Astrophysics* (eds Burns, M. L., Harding, A. K. & Ramaty, R.) 267–272 (American Institute of Physics, New York, 1983).
16. Sunyaev, R. A. & Trumper, J. *Nature* **179**, 506–508 (1979).
17. Shapiro, S., Lightman, A. & Eardley, D. *Astrophys. J.* **204**, 187–199 (1976).
18. Leventhal, M., MacCallum, C. J. & Stang, P. D. *Astrophys. J.* **225**, L11–L14 (1978).
19. Riegler, G. R. *et al. Astrophys. J.* **248**, L13–L16 (1981).
20. Lingenfelter, R. E. & Ramaty, R. in *The Galactic Centre* (eds Riegler, G. R. & Blandford, R. D.) 148–159 (American Institute of Physics, New York, 1982).
21. Bally, J., Stark, A. A., Wilson, R. W. & Henkel, C. *Astrophys. J. suppl. Ser.* **65**, 13–82 (1987).
22. Bally, J., Stark, A. A., Wilson, R. W. & Henkel, C. *Astrophys. J.* **324**, 223–247 (1988).
23. Bhat, Ch. L., Issa, M. R., Houston, B. P., Mayer, C. J. & Wolfendale, A. W. *Nature* **314**, 511–514 (1985).
24. Bloeman, J. B. G. M. *et al. Astr. Astrophys.* **154**, 25–30 (1986).
25. Bondi, H. & Hoyle, F. *Mon. Not. R. astr. Soc.* **104**, 273–279 (1944).
26. Lang, K. R. in *Astrophysical Formulae* 311–313 (Springer, Berlin, 1980).
27. Spitzer, L. *Physical Processes in the Interstellar Medium* (Wiley, New York 1978).

ACKNOWLEDGEMENTS. We thank G. Skinner, L. Bildsten and R. W. Wilson for useful comments.

A filamentary radio source near the Galactic Centre

A. D. Gray*, L. E. Cram*, R. D. Ekers† & W. M. Goss‡

* School of Physics, University of Sydney, New South Wales 2006, Australia

† Australia Telescope National Facility, PO Box 76, Epping, New South Wales 2121, Australia

‡ National Radio Astronomy Observatory, PO Box 0, Socorro 87801, New Mexico, USA

NEAR the Galactic Centre are several filamentary radio structures that have become known as ‘threads’¹ and which are thought to signify either magnetic loops² or electric current paths³ between active regions and the galactic magnetic field. Using the Molonglo Observatory Synthesis Telescope, we have found a new elongated radio source, G359.1–0.2, within 1° of the Galactic Centre. Observations with the Australia Telescope Compact Array and the Very Large Array show it to be filamentary in nature, extending more than 20 arcmin in length but only 10 arcsec wide. This object is spectrally similar to the previously known threads, but unlike them has kinks along its length, and does not appear to be obviously connected with any active region that might be responsible for its generation.

The discovery observation of the source described here was made as part of a survey of the Galactic Centre carried out using the MOST^{4,5} covering the $10^\circ \times 5^\circ$ ($l \times b$) region around the Galactic Centre. Operating in the radio continuum at 843 MHz with a bandwidth of 3 MHz, the MOST uses earth-rotation synthesis to image, in right-hand circular polarization, fields of dimension $70' \times 70'$ cosec (δ) ($\alpha \times \delta$) with an angular resolution of $43'' \times 43''$ cosec (δ) and a theoretical sensitivity of 0.5 mJy per beam after 12 h. The MOST does not respond to correlations on baselines shorter than 15.2 m (42.9λ), and hence does not image components in the ratio emission having spatial scales larger than $\sim 30'$. The very bright but smoothly varying ridge of nonthermal emission from the Galactic plane⁶ is therefore not visible in MOST images. The flux density scale of the

MOST is established by a group of southern flux density standards for which 1934–638 is set to 19.60 Jy.

Figure 1 shows G359.1–0.2 and the other complex sources in its neighbourhood. This 12-h MOST observation is limited to a sensitivity of about 2 mJy per beam by sidelobes from Sgr A. The brightness distribution along the axis of the object varies smoothly and continuously, peaking near the centre and tapering towards each end. There is no trace of diffuse emission towards either side of the object, suggesting that it is tube- or thread-like rather than a sheet or shell viewed edge-on. The object is kinked in two places, lending it a sinuous appearance which has led us to dub it ‘the Snake’.

Subsequent to its discovery, the Snake was observed with five antennas of the ATCA⁸ at a frequency of 4,790 MHz ($\lambda = 6.3$ cm) with a bandwidth of 128 MHz. The object was observed for 12 h on each of three pointing centres spaced along its length, as it is much longer than the diameter of the primary beam at this frequency. These observations were then mosaiced and simultaneously deconvolved using a maximum entropy technique (AIPS task VTESS⁹). Figure 2 shows the resulting image. A 6-h VLA¹⁰ observation in the B/C configuration at 1,446 MHz ($\lambda = 20.7$ cm) with a 12.5-MHz bandwidth was also obtained and is shown in Fig. 3. The spectral information obtained from these images shows that the Snake is nonthermal with a spectral index of $\alpha = -0.5 \pm 0.1$ ($S_\nu \propto \nu^\alpha$). The high spatial resolution of these images ($\sim 10''$) shows the Snake to be barely resolved across its shorter dimension.

All three images show that the southern end of the Snake crosses the shell supernova remnant SNR G359.1–0.5^{11,12} (resolved out in Fig. 2), but our current observations do not clearly show how far beyond the edge of the shell the object extends. Certainly the Snake can be traced for several arcmin within the outer boundary of the SNR shell and it may extend beyond the region shown, but examination of a field adjacent to that shown in Fig. 1 shows no evidence for a re-emergence from the other (eastern) side. In Fig. 1 the Snake becomes indistinct where its surface brightness approaches that of the SNR, but in Figs 2 and 3 the primary beam attenuation reduces the sensitivity towards the edge of the field. There is a suggestion in Fig. 3, however, that the Snake changes direction as it crosses



FIG. 1 The CLEANed MOST 843 MHz (35.6 cm) image. The Snake is the object running diagonally from upper-right (north-west) to lower-left (south-east). The shell-like source at the southern end is part of SNR G359.1 – 0.5 and at the northern end lies the H II region G359.2 – 0.0. Just off the displayed field to the north lies Sgr C and bright extended source to the east of the Snake is another H II region (FIR-27'). The resolution is $43'' \times 87''$ ($PA = -0.42^\circ$) and the r.m.s. noise is 2 mJy per beam. Coordinates in this and other figures are for epoch and equinox J2000.0.

or passes through the knotty emission of the SNR G359.1 – 0.5. Although the Snake does not seem to be similar to jet-like protrusions from SNRs¹³, it is interesting to note that this SNR is believed to be associated with the jet-like source G359.23 – 0.82 (ref. 14), which protrudes from the SNR shell to the east. Our images show that the body of the Snake is not well aligned with G359.23 – 0.82. It is thus unlikely that the two objects form a single continuous structure, nor is it likely that they originate from a common source (such as a pulsar) at or near the centre of the SNR shell. At its northern end the Snake fades smoothly, disappearing just before the triple-core of the H II region G359.2 – 0.0 (ref. 15), but extrapolation along a curve dictated by the visible portion of the Snake suggests that the H II region is not associated with the termination of the Snake.

Although the line-of-sight to the Snake passes near the Galactic Centre, the actual distance to the object is unknown. If it does lie at a distance similar to that of the Galactic Centre (8.5 kpc) then its angular size of $20'$ corresponds to a projected length of 50 pc (1.5×10^{20} cm) and its width of $10''$ to 0.5 pc (1.5×10^{18} cm). The morphology of the Snake is reminiscent of the class of objects known as 'threads'¹, although there are several significant differences in this new object. Firstly, the Snake lies outside the so-called Ω -lobe¹⁶, in the interior of which all other known threads exist¹⁷. Second, it lies farther from Sgr A than any other similar structures and is the only known thread or filament that extends into southern Galactic latitudes, aside from the cluster of filaments in the Sgr A 'Arc'¹⁸. Finally, and perhaps most importantly, the Snake is not monotonically curved and shows a prominent kink or dislocation at $\alpha_{J2000} = 17$ h 44 min 20 s, $\delta_{J2000} = -29^\circ 46' 35''$ and a lesser kink at $\alpha_{J2000} = 17$ h 44 min 34 s, $\delta_{J2000} = -29^\circ 48' 57''$. The possibility that the kinks

are associated with a helical geometry, although unlikely, cannot be ruled out by our observations. At the vertex of the former (major) kink a weak, unresolved source may be discerned in Figs 2 and 3. It is not clear whether this source is associated with the Snake, but its location at the vertex is suggestive of some form of interaction. In the VLA image (Fig. 3) it appears that at least to the south of the minor kink, and perhaps north of it as well, the Snake may in fact be composed of two adjacent threads. This behaviour is not unprecedented; the object G359.54 + 0.18 shows similar complex structure¹⁹. This latter object is believed to be the result of an interaction of the Galactic magnetic field with a molecular cloud. However, no molecular cloud is known to exist in the vicinity of the Snake.

Despite the differences between the Snake and other threads, the morphological similarities strongly suggest that they are manifestations of a common phenomenon. This hypothesis is strengthened by the fact that the spectral index of the Snake is consistent with figures derived for the threads¹⁷. The possibility that the underlying mechanism is electrodynamic in nature has been raised¹ and developed into comprehensive models^{2,3}. Our observations of the Snake provide useful tests of the assumptions underlying these models and of their predictions. One argument² is that the electrodynamic activity of the Galactic Centre occurs within, and on the surface of, the Ω -lobe or 'Galactic Centre Lobe' (GCL), modelled as a barrel-shaped region centred on Sgr A with its axis normal to the plane of the Galaxy, a radius of 30–50 pc and a height of ~ 100 pc. The activity is supposed to originate in a small (≤ 1 pc) generator of 'magnetic loops' lying deep within the GCL. Isolated threads are interpreted as portions of these magnetic loops that have connected with the GCL and expanded outwards from the plane over a scale of ~ 100 pc. This theory offers no explanation for the threads that

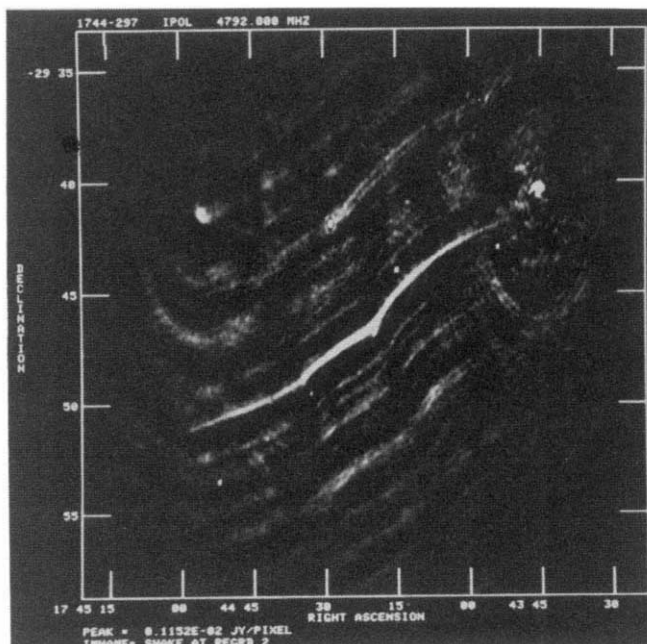


FIG. 2 The ATCA 6-cm observations. The image presented here is a mosaic comprised of a set of six 12-h observations, two on each of three separate pointing centres located at approximately $6'$ intervals along the length of the Snake. This image was deconvolved using a maximum entropy algorithm. The resolution is $7'' \times 12''$ ($PA = 6.23^\circ$) and the r.m.s. noise is 0.3 mJy per beam. Because of the relatively sparse coverage of the (u, v) -plane the maximum-entropy deconvolution was not entirely successful, so there are distinct sidelobes visible running parallel to the source in this image.

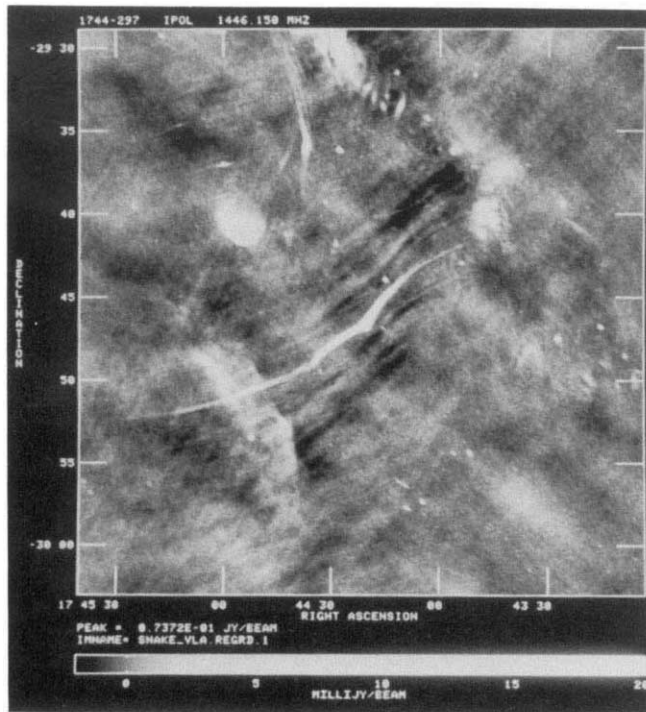


FIG. 3 A 6-h observation at 20 cm using the VLA in its B/C configuration. This CLEANed image has a resolution of $10'' \times 13''$ ($PA = 53.71^\circ$) and an r.m.s. noise of 0.4 mJy per beam.

make up the Snake, which transect the plane but lie outside the GCL. The discovery of the Snake thus violates the prediction that threads should lie between the generator and the GCL; whether this is fatal to the entire theory remains to be seen.

Another argument³ claims that 'threads' may delineate electric current paths that have been pulled away from active regions (such as the continuum arc) by magnetic tension. The possible existence of filamentation in the Snake and the slow, smooth curvature of its components may be consistent with this picture. If the Snake has been pulled a distance ~ 100 pc from a birth-place near the Galactic Centre and has lived no longer than the ohmic time scale of the generator ($\sim 10^4$ yr)³, its transverse proper motion could be as large as 0.2 arcs yr⁻¹. This fact, taken together with the suggestion that rapid (~ 2 yr) 'flares' might occur in the filaments, suggests that occasional repeated observation of the Snake would be warranted. Given the initial attraction of this concept it would be particularly useful to have more detailed theoretical models of the dynamical behaviour of current paths and of their propensity to sustain nonthermal radiation. $\square$

Received 5 June; accepted 3 July 1991.

1. Morris, M. & Yusef-Zadeh, F. *Astr. J.* **90**, 2511–2513 (1985).
2. Heyvaerts, J., Norman, C. & Pudritz, R. E. *Astrophys. J.* **330**, 718–736 (1988).
3. Benford, G. *Astrophys. J.* **333**, 735–742 (1988).
4. Mills, B. Y. *Proc. astr. Soc. Aust.* **4**, 156–159 (1981).
5. Whiteoak, J. B. Z., Large, M. I., Cram, L. E. & Piestrzynski, B. *Proc. astr. Soc. Aust.* **8**, 176–179 (1989).
6. Downes, D. & Maxwell, A. *Astrophys. J.* **146**, 653–665 (1966).
7. Odenwald, S. F. in *The Center of the Galaxy*, Proceedings of IAU Symposium 136 (ed. Morris, M.) 205–208 (Kluwer Academic, 1989).
8. Norris, R. P. *Sky & Telescope* **76**, 615–620 (1988).
9. Cornwell, T. J. in *Synthesis Imaging in Radio Astronomy*, (eds Perley, R. A., Schwab, F. R. & Bridle, A. H.) Astronomical Society of the Pacific Conference Series **6**, 277–286 (1989).
10. Napier, P. J., Thompson, R. & Ekers, R. D. *Proc. IEEE* **11**, 1295–1320 (1983).
11. Downes, D., Goss, W. M., Schwarz, U. & Wouterloot, J. G. A. *Astr. Astrophys. Suppl.* **35**, 1–22 (1984).
12. Reich, W. & Fürst, E. *Astr. Astrophys. Suppl.* **57**, 165–167 (1984).
13. Kesteven, M. J. et al. *Aust. J. Phys.* **40**, 855–859 (1987).
14. Yusef-Zadeh, F. & Bally, J. *Nature* **330**, 455–458 (1987).

15. Caswell, J. L. & Haynes, R. F. *Astr. Astrophys.* **171**, 261–276 (1987).
16. Sofue, Y. & Handa, T. *Nature* **310**, 568–569 (1984).
17. Anantharamaiah, K. R., Pedlar, A., Ekers, R. D. & Goss, W. M. *Mon. Not. R. astr. Soc.* **249**, 262–281 (1991).
18. Yusef-Zadeh, F., Morris, M. & Chance, D. *Nature* **310**, 557–561 (1984).
19. Bally, J. & Yusef-Zadeh, F. *Astrophys. J.* **336**, 173–177 (1989).

ACKNOWLEDGEMENTS. The MOST is operated with funds from the Australian Research Council and the Science Foundation for Physics. The Australia Telescope is operated in association with the Division of Radiophysics by the CSIRO. The VLA of the NRAO is operated by Associated Universities Inc. under cooperative agreement with the NSF.

Direct measurement of colloidal forces using an atomic force microscope

William A. Ducker*, Tim J. Senden
& Richard M. Pashley†

Department of Applied Mathematics, Research School of Physical Sciences, Australian National University, GPO Box 4, Canberra City, Australian Capital Territory, Australia

† Department of Chemistry, Faculty of Science, Australian National University, GPO Box 4, Canberra City, Australian Capital Territory, Australia

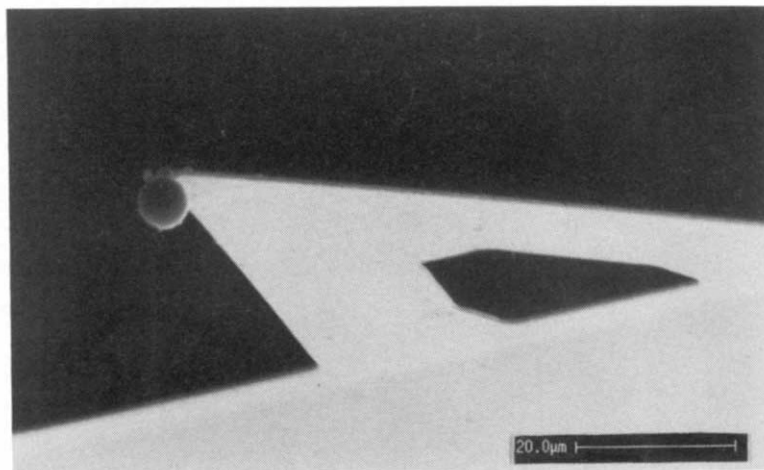
THE forces between colloidal particles dominate the behaviour of a great variety of materials, including paints, paper, soil, clays and (in some circumstances) cells. Here we describe the use of the atomic force microscope to measure directly the force between a planar surface and an individual colloid particle. The particle, a silica sphere of radius $3.5 \mu\text{m}$, was attached to the force sensor in the microscope and the force between the particle and the surface was measured in solutions of sodium chloride. The measurements are consistent with the double-layer theory^{1,2} of colloidal forces, although at very short distances there are deviations that may be attributed to hydration forces^{3–6} or surface roughness, and with previous studies on macroscopic systems^{4–6}. Similar measurements should be possible for a wide range of the particulate and fibrous materials that are often encountered in industrial contexts, provided that they can be attached to the microscope probe.

Fifty years ago, Derjaguin, Landau, Verwey and Overbeek¹ developed a quantitative theory (DLVO theory) to describe the stability of colloid particles in aqueous solution. According to this theory, colloid stability is dependent on the balance between two opposing forces: an attractive van der Waals force, and a repulsive electrostatic force known as a double-layer force. During the past 20 years the main features of DLVO theory have been confirmed⁷ with considerable modifications to include liquid structure⁸, solvation⁹ and hydrophobic¹⁰ and other forces. Because of the difficulty in working with such small particles (micrometres to nanometres), experimental studies have used a limited range of macroscopic substrates. Alternatively, indirect methods such as light or neutron scattering¹¹ and sedimentation studies¹² have been used. Only total internal reflectance microscopy¹³ has come close to the goal of force measurements on individual particles. But recent advances in force measurement and distance control associated with the development of the scanning tunnelling microscope (STM)¹⁴ and atomic force microscope (AFM)¹⁵ now allow direct measurement of the forces on an individual colloid particle. Previous force measurements using AFMs have concentrated on measurements of van der Waals forces between a tip and sample in vapour^{16,17}, and little work has been done in liquid media^{18,19}. In the studies done so far, geometry and surface chemistry were poorly characterized and comparison with theory is difficult.

We have measured the force on a small silica sphere as it

* Present address: Department of Chemical and Nuclear Engineering, University of California, Santa Barbara, California 93106, USA.

FIG. 1 A scanning electron microscope image of a colloid probe of the type used in our experiments. The glass sphere, $3.5\text{-}\mu\text{m}$ radius, was glued with $\sim 10^{-15}$ l of resin (epikote 1004, Shell) to the free end of a V-shaped silicon nitride cantilever (Park Scientific) using a three-dimensional micro-translation stage. The deflection of the cantilever is determined by the change in position of a laser spot reflected from the free end onto a pair of photodiodes²³.



approaches a flat silica surface in aqueous NaCl solutions. Silica was selected both because of its common occurrence and because previous investigation using other techniques provides a basis for comparison. Spherical geometry was chosen for ease of comparison with predicted forces. We used a silica-glass sphere (Polysciences Inc.) and a flat silicon wafer (Okemtic, Finland) oxidized to a depth of 30 nm. Both the sphere and the flat were cleaned immediately before use by exposure to a water plasma (18 MHz, 25 W for 2 min, partial pressure of water, $P_{\text{H}_2\text{O}} = 0.02$ torr and $P_{\text{Ar}} = 0.03$ torr). This resulted in an advancing water contact angle of less than 10° , and ensured a high density of surface silanol groups. The surface roughness of these substrates was measured using an AFM. In a typical $(600\text{ nm})^2$ area the standard deviation from mean height was 0.3 nm for the sphere and 0.2 nm for the flat. The highest asperities were 1.6 and 0.7 nm, respectively.

Force-distance relationships were measured with a Nanoscope II AFM (Digital Instruments). This instrument produces a signal proportional to the deflection of a cantilever as a function of sample displacement. In normal operation this is a measure of the force on a sharp tip. We have modified the usual AFM probe, however, by replacing the sharp tip with a colloid particle (Fig. 1). To convert the data on deflection against sample displacement to a curve of force as a function of tip-sample displacement, it is necessary to define zeros of both force and distance, and to convert the deflection signal to force. The zero of deflection was chosen where the deflection was constant (with the particle and flat far apart), but the zero of distance is a more difficult choice. As the sample is driven toward the sphere, the cantilever deflects, and this is registered by a photodiode. At some point, the output of the diode becomes a linear function of displacement of the sample, indicating that the motion of the tip and sample are coupled. In this regime of constant compliance, the particle is 'in contact' with the surface, and we have defined this separation to be the zero of distance. The relationship between displacement and diode response in this region was independent of NaCl concentration, and was used to convert the diode response to the effective displacement of the cantilever. This relationship was also used to determine changes in the sphere-flat separation when the two were not in contact, and to calculate the surface force. An overestimate of surface force will be obtained if the cantilever is not significantly more compliant than the sphere, the flat surface and the components connecting them. For compliant substrates, this problem can be overcome by independently calibrating the cantilever deflection.

Interaction forces measured for a silica-glass sphere immersed in a range of aqueous NaCl solutions are shown in Fig. 2. As predicted by DLVO theory, the measured forces decay exponen-

tially at large separation, and both the decay length and electrostatic potential decrease with concentration. The predicted force can be fitted to the measured force at separations greater than about one decay length by assuming a surface potential for the silica. The potentials fitted from the data in Fig. 2 are, as expected, somewhat lower than those obtained from streaming

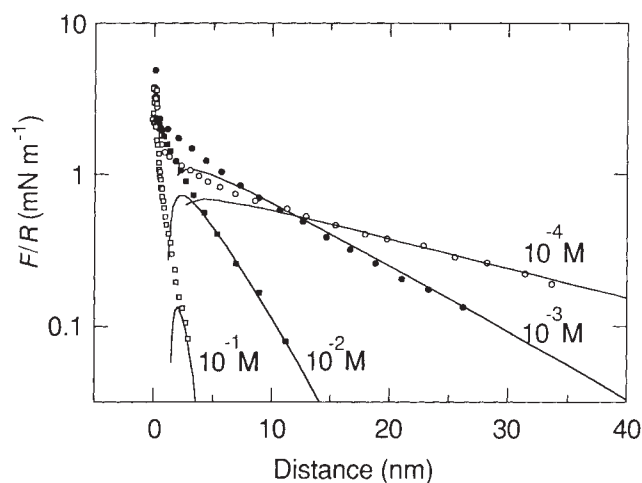


FIG. 2 The force, F , as a function of distance, D , for a silica probe of radius $R = 3.5\text{ }\mu\text{m}$. The force has been normalized by the sphere radius because $F/2\pi R$ is equal to the energy per unit area between two equivalent flat surfaces (according to the Derjaguin approximation²⁴). These forces were measured using a cantilever of stiffness 0.58 N m^{-1} and an approach velocity of less than 200 nm s^{-1} . Doubling the approach rate produced no change in the measured forces, indicating that hydrodynamic forces were insignificant. The zero of distance was chosen to occur where an additional applied force resulted in no further reduction in sphere-flat displacement. The errors in an individual measurement are $\sim \pm 0.2\text{ nN}$ in force, $\pm 0.3\text{ nm}$ in distance. There is also a total of $\pm 5\%$ in systematic errors due to measurement of the particle radius and the spring constant, and determination of the regime of constant compliance. For clarity each data point shown is the average of 10–15 original data points. The solid lines show forces calculated from DLVO theory. These forces are the sum of a nonretarded van der Waals attraction ($F_{\text{vdw}}/R = A/6D^2$, where A is the Hamaker constant, equal to $0.85 \times 10^{-20}\text{ J}$ (ref. 25) in this case) and a repulsive electrical double-layer force calculated from the Poisson-Boltzmann equation using an exact numerical solution²⁶. For different molar concentrations of NaCl, the following values of surface potential, Ψ_0 , and decay length, κ^{-1} , were used to fit the curves: 10^{-1} M : $\Psi_0 = 61\text{ mV}$, $\kappa^{-1} = 1.1\text{ nm}$; 10^{-2} M : $\Psi_0 = 53\text{ mV}$, $\kappa^{-1} = 3.2\text{ nm}$; 10^{-3} M : $\Psi_0 = 34\text{ mV}$, $\kappa^{-1} = 9.1\text{ nm}$; 10^{-4} M : $\Psi_0 = 21\text{ mV}$, $\kappa^{-1} = 21\text{ nm}$. For these calculations, the origin of charge was taken to be at the point of closest approach of each surface.

potential measurements on silica capillaries²⁰, but somewhat higher than those obtained by Horn *et al.*⁴ from force measurements on large silica sheets. In the latter case, however, the silica showed an advancing contact angle of 45°, indicating a lower density of surface silanol groups than for the hydrophilic silica used in our experiments.

At smaller separations, the force predicted by DLVO theory depends on the choice of boundary condition for the surface charge. Our measurements lie between the limits of constant charge and constant potential, but for clarity, only the theoretical interaction at constant potential has been shown in Fig. 2. At very small separations (2–3 nm) DLVO theory predicts that the attractive van der Waals component will exceed the double-layer force, but the measured force is greater than even the limit of constant charge. This effect has been seen previously, and has been attributed to hydration forces^{3–6}. In our case, however, the roughness of the substrates complicates analysis of short-range forces.

Variation between the results of surface-potential measurements for different samples illustrates the advantage of force measurements on the actual particles of interest. Minor differences in surface composition or morphology can have a large effect on particle–particle interactions, so it is an advantage to be able to investigate a particle rather than a model system. Although the force measurements presented here agree well with those obtained from macroscopic systems, there may be differences in the interaction forces when pairs of macroscopic or colloid particles interact. For example, because hydrodynamic forces (on a particle approaching a surface) scale with the square of particle radius²¹, it is much easier, and of more relevance, to test the influence of relaxation times for surface forces on small particles than on macroscopic systems.

The leading technique for the measurement of surface forces is currently the apparatus of Israelachvili and Adams^{7,22}. In that technique the force is measured between a limited number of macroscopic surfaces in crossed cylinder geometry. In contrast, use of a colloid probe allows investigation of particulate and fibrous materials. Particles in the mineral processing and ceramics industries, colloid polymer industries and in biological systems can now be studied directly. □

Dimension of weather and climate attractors

Edward N. Lorenz*

Institute for Nonlinear Science, University of California at San Diego, La Jolla, California 92093, USA

* Present address: Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

A PROCEDURE for estimating the correlation dimension of the attractor of a dynamical system¹ has been applied to a number of data sets that are representative of weather or climate variations. Reported values of the attractor dimension have typically fallen between 3.0 and 8.0 (refs 2–9). Because the atmosphere is so complex, these values have seemed surprisingly low, and doubts as to their appropriateness have been expressed even by the originators of the method^{8,10,11}. Here I apply the procedure to 'data' generated by a mathematical system whose dimension can be evaluated by other means, and identify conditions, apparently satisfied by the studies that use real data, in which the procedure will yield systematic underestimates. It therefore seems unlikely that global weather or climate systems possess low-dimensional attractors.

The quantities selected for analysis have included sea-level pressures², upper-level geopotential heights^{3,4}, low-level vertical velocity components⁵, rainfall intensities⁶, sunshine durations², cyclone positions⁷, large-scale wave amplitudes², tree-ring widths⁸ and oxygen-isotope concentrations in deep-sea cores^{2,8,9}. The findings have raised hopes that suitably constructed models with relatively few variables might recapture the dynamics of the weather or the climate.

In a typical study one begins with an extensive sequence of values of a single scalar quantity $x(t)$, observed at equally spaced times t . The corresponding sequence of vectors $[x(t), y(t), \dots]$, where $y(t), \dots$ are other scalar quantities coupled to $x(t)$, constitutes a sample of a dynamical system, but in general the values of $y(t), \dots$ are not known. One therefore attempts to reconstruct the system by choosing a time lag τ and an embedding dimension

Received 19 March; accepted 23 July 1991.

- Derjaguin, B. & Landau, L. *Acta Physicochem.* **14**, 633 (1941).
- Verwey, E. G. W. & Overbeck, J. J. G. *Theory of the Stability of Lyophobic Colloids* (Elsevier, Amsterdam, 1948).
- Proc. Nobel Conf. Hydration Forces and Molecular Aspects of Solvation Chem. Scr. **25**, 3–31 (1985).
- Horn, R. G., Smith, D. T. & Haller, W. *Chem. Phys. Lett.* **162**, 404–408 (1989).
- Rabinovich, I., Derjaguin, B. V. & Churaev, N. V. *Adv. Colloid Interf. Sci.* **16**, 63–78 (1982).
- Peschel, G., Belouschek, P., Muller, M. M., Muller, M. R. & Konig, R. *Colloid Polym. Sci.* **260**, 444–451 (1982).
- Israelachvili, J. N. & Adams, G. E. *JCS Faraday Trans. I* **74**, 975–1001 (1978).
- Horn, R. G. & Israelachvili, J. N. *Chem. Phys. Lett.* **71**, 192–194 (1980).
- Pashley, R. M. *J. Colloid Interf. Sci.* **83**, 531–546 (1981).
- Pashley, R. M., McGuiggan, P. M., Ninham, B. W. & Evans, D. F. *Science* **229**, 1088–1089 (1985).
- Ottewill, R. H. *Concentrated Dispersions in Colloid Dispersions* Ch. 9 (ed. Goodwin, J. W.) (Royal Society of Chemistry, London, 1982).
- Elimelech, M. *JCS Faraday Trans. I* **86**, 1623–1624 (1990).
- Zheng, X. & Yoon, R. *J. Colloid Interf. Sci.* **134**, 427–434 (1990).
- Brown, M. A. & Staples, E. J. *Langmuir* **6**, 1260–1265 (1990).
- Prieve, D. C. & Freij, N. A. *Langmuir* **6**, 396–403 (1990).
- Binnig, G. & Rohrer, H. *Helv. Phys. Acta* **55**, 726–735 (1982).
- Binnig, G., Quate, C. F. & Gerber, C. *Phys. Rev. Lett.* **56**, 930–933 (1986).
- Martin, Y., Williams, C. & Wickramasinghe, H. *J. appl. Phys.* **61**, 4223–4229 (1987).
- Burnham, N. A. & Colten, R. J. *J. Vac. Sci. Technol. A7*, 2906–2913 (1989).
- Ducker, W. A. & Cook, R. F. *Appl. Phys. Lett.* **56**, 2048–2410 (1990).
- Weisenhorn, A. L., Hansma, P. K., Albrecht, T. R. & Quate, C. F. *Appl. Phys. Lett.* **54**, 2691–2653 (1989).
- Wiese, G. R., James, R. O. & Healy, T. W. *Disc. Faraday Soc.* **52**, 302–311 (1975).
- Chan, D. Y. C. & Horn, R. G. *J. chem. Phys.* **93**, 5311–5324 (1990).
- Parker, J. L., Christenson, H. K. & Ninham, B. W. *Rev. sci. Instrum.* **60**, 3135–3138 (1989).
- Meyer, G. & Amer, N. M. *Appl. Phys. Lett.* **53**, 1045–1047 (1988).
- Derjaguin, B. V. *Kolloid. Zh.* **69**, 155–164 (1934).
- Hunter, R. J. *Foundations of Colloid Science*, 222 (Clarendon, Oxford, 1987).
- Chan, D. Y. C., Pashley, R. M. & White, L. R. *J. Coll. Inter. Sci.* **77**, 283 (1980).

ACKNOWLEDGEMENTS. We thank B. W. Ninham for support, J. Parker for the use of his plasma reactor, T. Nylander for providing the oxidized wafer, and C. Drummond and T. W. Healy for discussions.

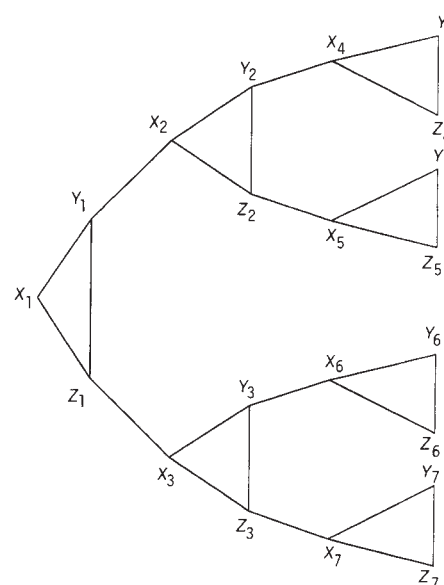
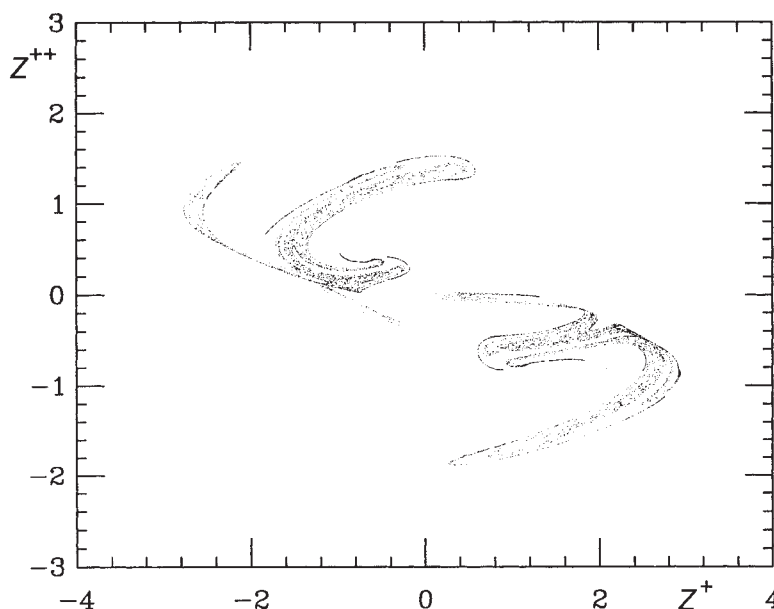


FIG. 1 Coupling scheme for variables in equation (1) with $L=3$. Line segments forming sides of triangles represent strong nonlinear coupling between variables with like subscripts; segments connecting triangles represent weak or strong linear coupling between variables with different subscripts.

FIG. 2 Poincaré section of reconstructed attractor of equation (1) with $L=1$, obtained by plotting $Z^{++}=Z_1(t'+2\tau)$ against $Z^+=Z_1(t'+\tau)$ for times when $Z_1(t')=0$.



n , and forming the vectors $[x(t), x(t+\tau), \dots, x(t+n\tau-\tau)]$ for N times t spaced at intervals τ' ; sometimes $\tau'=\tau$. Finally one determines by direct count, for various values of r , the number $m(r)$ of pairs of vectors whose vector difference is less than r in magnitude. If, as r approaches zero, $m(r)$ varies as $r^{d(n)}$, the estimated correlation dimension of the attractor of the reconstructed system is $d(n)$. If, as n is successively increased, $d(n)$ approaches a limit d_c , the estimated correlation dimension of the attractor of the original system is d_c .

A number of factors that might produce misleading results have been noted^{12,13}; some of them can be overcome by simple modifications. Here I propose another explanation for the seemingly low estimates.

A key phrase in defining d_c is 'as r approaches zero.' As pointed out by Ruelle¹⁴, unless N is very large the smallest value of r for which $m(r)$ is large enough to be statistically meaningful may not be particularly close to zero. If, for example, $N=4,000$, and if in reality $d_c=8.0$, $m(r)$ can decrease all the way from its maximum, about 8×10^6 , to unity while r decreases by a factor of less than eight.

In concluding that d_c is small, then, the investigators have implicitly assumed that the fine structure of the attractor would,

on magnification, resemble the coarse structure, so that, if N were made large enough, the decrease of $m(r)$ with r when r is very small would be like that when r is fairly large. Although this is demonstrably the case for some simple systems, it need not be so for more intricate ones, including some in which certain subsets of the variables are only weakly coupled to others. In the atmosphere much of the coupling is weak and indirect. If convective-cloud circulations over London and Washington, for example, can affect each other at all, they can do so only by being coupled to the general weather patterns in their own neighbourhoods, which may in turn be coupled to the same globe-encircling westerly wind current. I propose that if the variable selected for analysis is strongly coupled to only a few variables of the system, the estimated value of d_c , if N is only moderately large, will be considerably less than the dimension as determined by other standard methods, such as the Kaplan-Yorke¹⁵ formula.

I have arrived at this hypothesis after studying special cases of a model with $(3 \times (2^L - 1))$ variables, with $L=1$ or $L=3$, where the Kaplan-Yorke dimension d' may be readily estimated. The model was formed by taking $2^L - 1$ copies of a three-variable model whose properties I had previously studied in detail¹⁶,

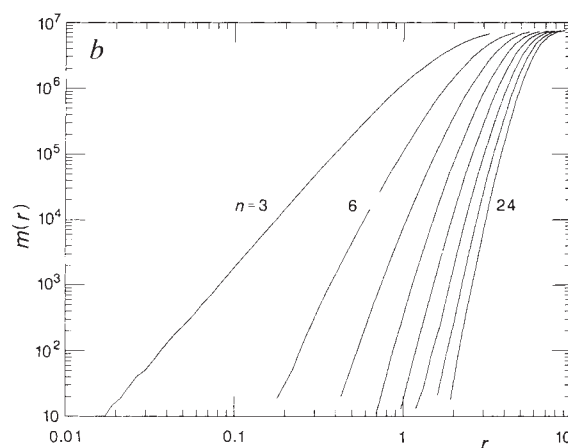
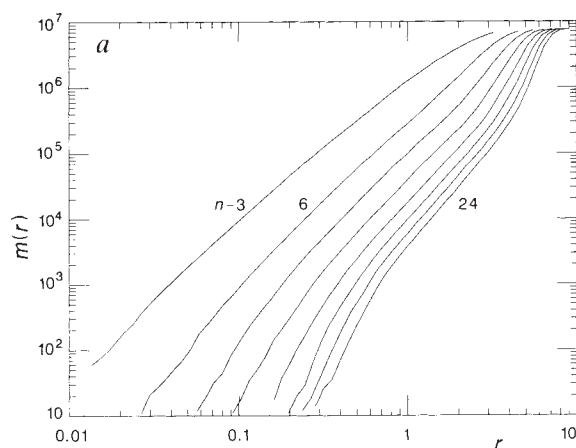


FIG. 3 *a*, Curves of $m(r)$ against r for embedding dimensions 3, 6, ..., 24, obtained by applying GP procedure to equation (1) with $L=3$ and $c=0.1$ and Z_7 as the selected variable. Points plotted for values of $\log_{10} r$ at

intervals of 0.05 have been connected by straight line segments. *b*, The same, except that $Z^*=Z_1 + \dots + Z_7$ is the selected variable.

and introducing linear couplings, as indicated schematically in Fig. 1. The equations of the new model are

$$dX_j/dt = -q_j(Y_j^2 + Z_j^2) - a_j(X_j - F_j) + U_j \quad (1a)$$

$$dY_j/dt = X_j(q_j Y_j - b_j Z_j) - p_j(Y_j - G_j) + V_j \quad (1b)$$

$$dZ_j/dt = X_j(b_j Y_j + q_j Z_j) - p_j Z_j + W_j \quad (1c)$$

for $j = 1, \dots, 2^L - 1$, where

$$U_{2j} = -cp_j Y_j \quad (2a)$$

$$U_{2j+1} = -cp_j Z_j \quad (2b)$$

$$V_j = cp_j(X_{2j} - h_{2j}) \quad (2c)$$

$$W_j = cp_j(X_{2j+1} - h_{2j+1}) \quad (2d)$$

for $j = 1, \dots, (2^{L-1} - 1)$.

In these calculations I have let $p_j = 1.0$, $q_j = 1.0$, $a_j = 0.25$, $b_j = 4.0$, $F_j = 8.0$, $G_j = 1.0$ and $h_j = 1.0$ for all values of j , while letting the coupling coefficient c be 0.1 for 'weak' and 1.0 for 'strong' coupling. I have used the 'standard' fourth-order Runge-Kutta procedure with a time step of 0.025 units in all numerical

integrations. In each application of the Grassberger-Procaccia (GP) procedure I have let τ and τ' equal 8 time steps, and, after generating a sequence of values of a selected variable, I have let $N = 4,000$ for various values of n . This value is typical of many of the real-world studies (although N is considerably larger in some).

For $L = 1$ the new model reduces to the three-variable model. With the indicated values of the parameters it behaves chaotically, with $d' = 2.4$. Figure 2 shows a Poincaré section of a 'reconstructed' attractor, obtained by using fourth-degree polynomial interpolation to identify a sequence of times t' at which $Z_1(t') = 0$, and then plotting values of $Z_1(t' + 2\tau)$ against values of $Z_1(t' + \tau)$. The GP procedure, with Z_1 as the selected variable, indicates a value of d_c slightly exceeding 2.0.

For $L = 3$ the indicated value of d' when $c = 0.1$ is 17.0. Figure 3a shows curves of $m(r)$ against r obtained when Z_7 is the selected variable. According to Fig. 1, Z_7 is very weakly coupled to most of the 21 variables, particularly Y_4 , Z_4 , Y_5 and Z_5 . The central portions of the right-hand curves suggest that d_c is near 4.0, whereas the lower portions suggest a value near 5.5. One might argue that the procedure has not converged to any answer, but there is certainly no suggestion of a value comparable to

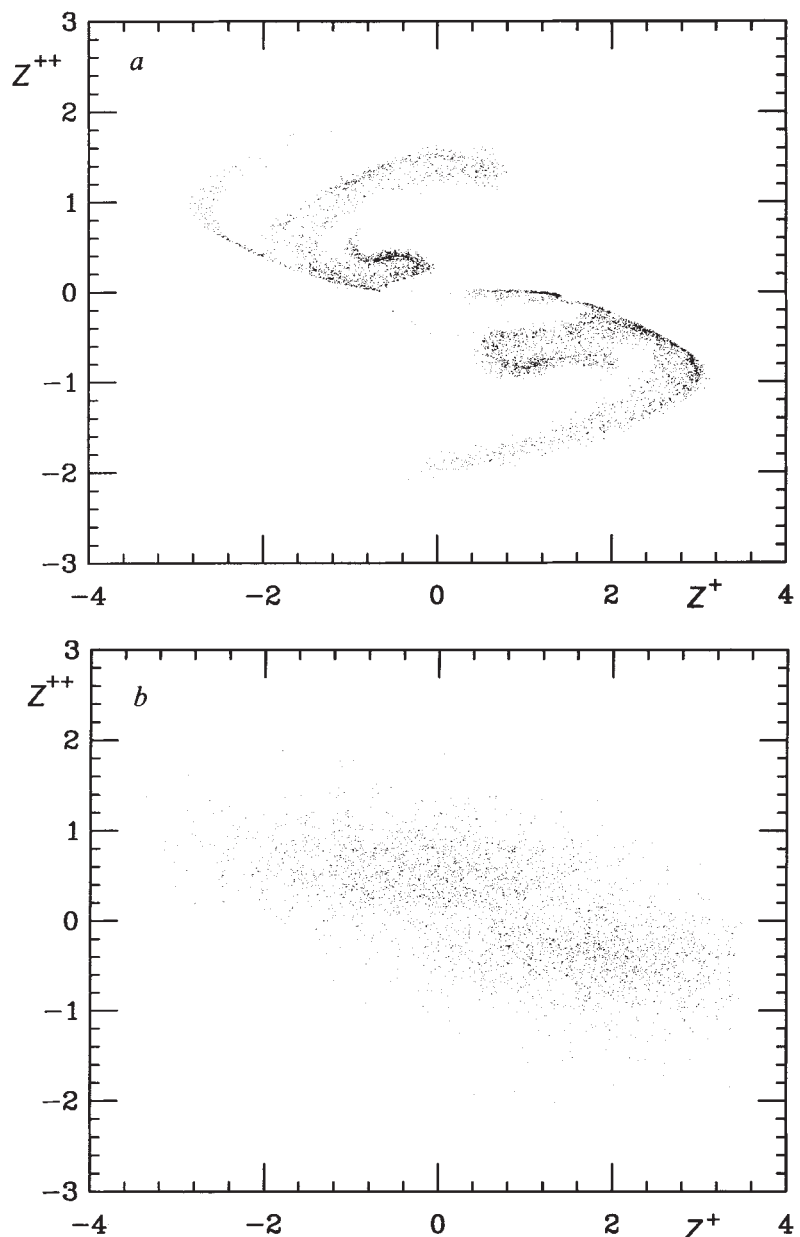


FIG. 4 a, Plane projection of Poincaré section of reconstructed attractor of equation (1) with $L = 3$ and $c = 0.1$, obtained by plotting $Z^{++} = Z_7(t' + 2\tau)$ against $Z^+ = Z_7(t' + \tau)$ for times when $Z_7(t') = 0$. b, The same, except that $Z^* = Z_1 + \dots + Z_7$ is used in place of Z_7 .

d' . Presumably a high value of d_c would have been obtained with a very high value of N .

The model equations for $L=3$ could be rewritten with any 21 independent linear combinations of the original 21 variables as new variables, and the right-hand sides would still contain only quadratic, linear and constant terms, while d' and d_c should remain unaltered. In particular, two of the new variables could be Z_7 and the sum $Z^* = Z_1 + \dots + Z_7$. We might expect Z^* to be fairly strongly coupled to most of the remaining variables.

Figure 3b is constructed similarly to Fig. 3a, but with Z^* as the selected variable. There is little resemblance between the figures. The right-hand curves suggest a value of d_c near 15.0, and hence comparable to d' . Repeating the calculations with $c=1.0$ gives qualitatively similar although less extreme results; d' is ~ 17.8 , and d_c is estimated to be ~ 7.5 when Z_7 is the selected variable and ~ 15.0 when Z^* is selected. It thus appears, when N is not too large, that, first, different selected variables can yield different estimates of d_c , and, second, a suitably selected variable can sometimes yield a fairly good estimate.

Figures 4a and 4b are constructed similarly to Fig. 2, but for $L=3$, with values of Z_7 in Fig. 4a and Z^* in Fig. 4b. Each figure therefore presents a projection of a Poincaré section of a high-dimensional attractor on a plane. The remarkable feature is that at coarse scales Fig. 4a looks very much like Fig. 2, although at fine scales it is more like Fig. 4b. The implication is that if N is so small that the smallest value of r for which $m(r)$ is statistically meaningful is still in the coarse-scale range, the estimate of d_c for $L=3$ with Z_7 as the selected variable

ought to resemble the estimate for $L=1$ when Z_1 is selected more than the one for $L=3$ when Z^* is selected.

I therefore see no reason to believe that an extensive weather or climate system possesses a low-dimensional attractor. At the same time, I do not feel that most of the real-data studies are meaningless; they merely need to be reinterpreted. As suggested in one study⁵, the atmosphere might be viewed as a loosely coupled set of lower-dimensional subsystems. Perhaps the procedure, as practised, attempts to measure the dimension of a subsystem. □

Received 1 May; accepted 6 August 1991.

1. Grassberger, P. & Procaccia, I. *Phys. Rev. Lett.* **50**, 346–349 (1983); *Physica D* **3**, 189–208 (1983).
2. Fraedrich, K. *J. Atmos. Sci.* **43**, 419–432 (1986).
3. Essex, C., Lookman, T. & Nerenberg, M. A. H. *Nature* **326**, 64–66 (1987).
4. Keppenne, C. L. & Nicolis, C. *J. Atmos. Sci.* **46**, 2356–2370 (1989).
5. Tsonis, A. A. and Elsner, J. B. *Nature* **333**, 545–547 (1988); *Bull. Am. Met. Soc.* **70**, 14–23 (1989).
6. Sharifi, M. B., Georgakakos, K. P. & Rodriguez-Iturbe, I. *J. Atmos. Sci.* **47**, 888–893 (1990).
7. Fraedrich, K. & Leslie, L. M. Q. *J. R. Met. Soc.* **A115**, 79–91 (1989).
8. Grassberger, P. *Nature* **323**, 609–612 (1986).
9. Nicolis, C. & Nicolis, G. *Nature* **311**, 529–532 (1984); **326**, 523 (1987).
10. Grassberger, P. *Nature* **326**, 524 (1987).
11. Procaccia, I. *Nature* **333**, 498–499 (1988).
12. Osborne, A. R. & Provenzale, A. *Physica D* **35**, 357–381 (1989).
13. Theiler, J. *J. Opt. Soc. Am.* **A7**, 1055–1073 (1990).
14. Ruelle, D. *Proc. R. Soc. A* **A427**, 241–248 (1990).
15. Kaplan, J. L. & Yorke, J. A. *Lect. Not. Math.* **730**, 204–219 (1979).
16. Lorenz, E. N. *Tellus A* **36**, 98–110 (1984); **A42**, 378–389 (1990).

ACKNOWLEDGEMENTS. I thank H. Abarband, R. Brown and S. Islam for their suggestions and assistance. This work was sponsored by the University Research Initiative of the Defense Advanced Research Project Agency and the Climate Dynamics Program of the NSF.

Wintertime asymmetry of upper tropospheric water between the Northern and Southern Hemispheres

K. K. Kelly*, A. F. Tuck & T. Davies†

* National Oceanic and Atmospheric Administration Aeronomy Laboratory, 325 Broadway, Boulder, Colorado 80303-3328 USA

† ECMWF, Reading, Berkshire, UK

WATER vapour is an important greenhouse gas^{1–3} and yet its abundance in the upper troposphere is poorly known. Upper-tropospheric water vapour is particularly important despite its low mixing ratios, because it has large effects on the flux of infrared radiation near the tropopause². In addition, the distribution and supply of water vapour are central to cloud formation; the effects of cloud on the Earth's radiation budget are in turn central to understanding the climate response to increasing atmospheric concentrations of greenhouse gases. From airborne measurements of total water (vapour plus ice crystal)⁴ during the winters of 1987 in the Southern Hemisphere and of 1988–89 in the Northern Hemisphere, we find that the upper troposphere in middle, subpolar and high latitudes is a factor of 2–4 drier during austral winter than during boreal winter. As the lower-latitude air moves towards the pole in austral winter, it is forced to cool to lower temperatures than in the north—more of the water vapour therefore condenses to form ice crystals, which then precipitate, thereby removing moisture from the air mass. Clearly, climate models must be able to reproduce this asymmetry if their predictions are to be credible. We also note that the asymmetry in water vapour implies an asymmetry in the production rate of the hydroxyl radical, and hence in the tropospheric chemistry of each hemisphere, for example in the rate of methane loss⁵.

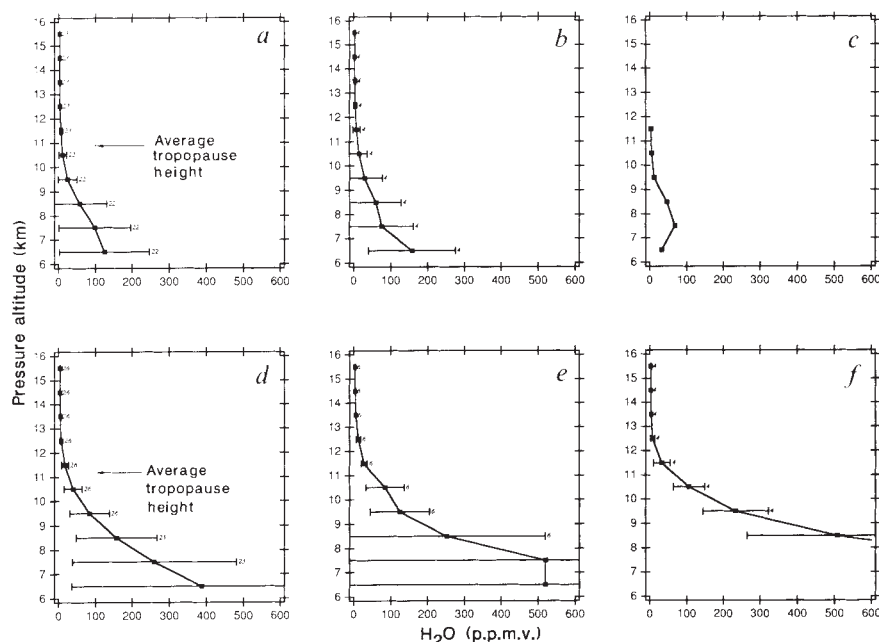
During the Airborne Antarctic Ozone Experiment (AAOE) and the Airborne Arctic Stratospheric Expedition (AASE), the

NASA ER-2 and DC-8 aircraft obtained many profiles of water in the upper troposphere above Punta Arenas (53° S, 71° W) and Stavanger (59° N, 6° E) during August to September 1987 and late December to February 1988/9 respectively. A lesser number of profiles was obtained at Puerto Montt (41° S, 73° W), Christchurch (44° S, 172° E), Moffett Field (37° N, 122° W) and Wallops Island (38° N, 75° W), all with the Lyman- α resonance fluorescence hygrometer⁴. The DC-8 aircraft completed more than 30 flights averaging >10 h duration on the two missions, mostly in the upper troposphere between 300 and 200 mbar. The average profile for each austral site and for each boreal site is shown in Fig. 1. For each set of three sites, the Northern Hemisphere contains a factor of 2–4 more water molecules in the upper troposphere than does the Southern Hemisphere. The relative frequencies of cloud near the tropopause at the sites have been discussed by Murphy *et al.*⁶. The difference in the upper-tropospheric water content is a striking extension downwards of the inter-hemispheric asymmetry in water observed in the stratosphere⁷ by the ER-2.

The large numbers of upper-tropospheric water observations obtained by the DC-8 in horizontal flight are shown in Fig. 2. Below the saturation curve the data refer to water vapour, above the curve they are the sum of vapour plus ice crystals. The asymmetry is present for both sets of data (clear air and clouds). The total water content of cloud is an important factor in cloud feedback as simulated by general circulation models^{8,9}. Note also that in clear air, the extreme dry population is made up entirely of austral points, whereas the extreme moist population consists of boreal points.

We examine air motions and temperatures associated with a very cold balloon ascent from the South Pole on 24 July 1987, using air parcel trajectories calculated from the analysis data from the European centre for medium-range weather forecasts (ECMWF). The temperature and pressure data were recorded from an ozonesonde flown by NOAA/CMDL¹⁰ and are shown in Fig. 3. It is clear from the ozonesonde that the tropopause is at ~ 180 mbar. This is not an unusual tropopause height over Antarctica in winter^{11–13}. The saturation mixing ratio for vapour over ice on this profile is less than 5 p.p.m.v. at 300 K, a potential

FIG. 1 The average profiles of total water from three Southern Hemisphere locations, August–September 1987, and from three Northern Hemisphere locations, December 1988–February 1989. The bars are standard deviations and represent atmospheric variability; the number to the right of each bar is the number of ER-2 flights contributing to the average at that altitude. *a*, Punta Arenas (53° S, 71° W); *b*, Puerto Montt (41° S, 73° W); *c*, Christchurch (44° S, 172° E). There was only one, DC-8, profile at Christchurch. *d*, Stavanger (59° N, 6° E), *e*, Moffett Field (37° N, 122° W); *f*, Wallops Island (38° N, 75° W).



temperature in the upper troposphere. The crests of upper-tropospheric ridges on 12-hourly geopotential height charts at 300 mbar frequently penetrate to 80° S and beyond during winter, particularly over West Antarctica. Comparison of forecast and analysed winds with those experienced by the DC-8 indicates that the real atmosphere has stronger meridional winds, and that ridges penetrate to higher latitudes, than in the meteorological analyses¹⁴. Schwerdtfeger¹⁵ gives frequencies of winds

greater than 10 m s^{-1} at the south Pole, and notes that the prevalence of this condition at 500 mbar for more than 50% of the time during July is incompatible with a polar symmetric vortex.

Three-dimensional five-day back trajectories from the South Pole in the upper troposphere, from 24 July 1987 back to 19 July, show that the air was transported from the 40°–50° S latitude belt over the Pacific and Atlantic Oceans, cooling by up to 40 K, decreasing in pressure by up to 250 mbar, and reaching saturation mixing ratios for water vapour over ice of less than 5 p.p.m.v. at 300 mbar in the ECMWF analysis. The 10-day forward trajectories from the South Pole show that some of this air had reached latitudes as low as 44° S by 3 August 1987. Pressure rises of up to 300 mbar accompanied this equatorward motion (Fig. 4). Saturation in the poleward-moving air masses is evident in satellite data for the day. Figure 5*a* shows

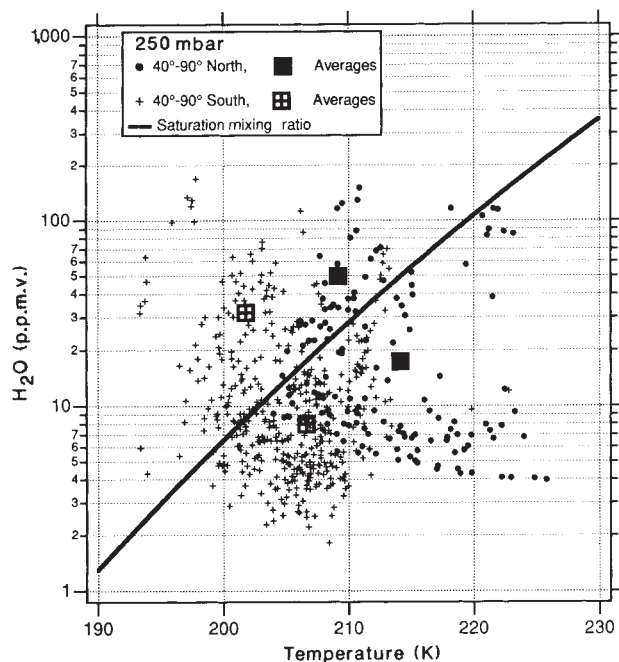


FIG. 2 Upper-tropospheric total water data from the DC-8. The instrument has 1 Hz response; each point represents a 300-s average in the pressure range 250 ± 25 mbar. Flights averaged over 10 h at 460 knots. Note that the colder Southern Hemisphere temperatures are in accord with climatology. The curve represents saturation mixing ratio with respect to ice; points above the line refer to cloud.

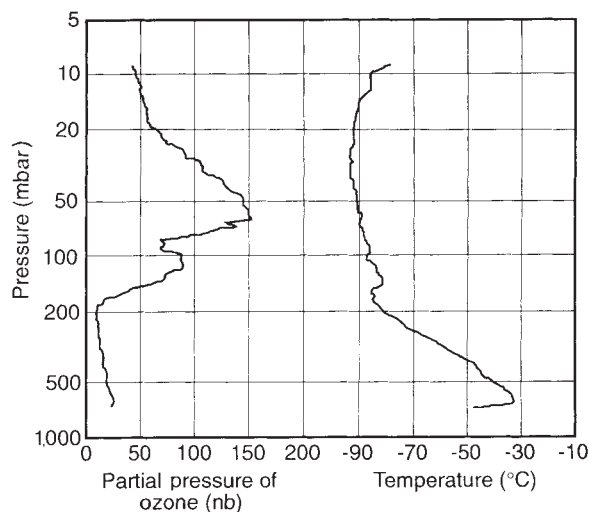


FIG. 3 Ozone and temperature from an ozonesonde ascent above the South Pole, 24 July 1987. The tropopause is at ~180 mbar, and saturation vapour pressure over ice is less than 5 p.p.m.v. above ~300 mbar.

infrared cloud images from orbiter F8 of the US Defense Meteorological Satellite Project; there is a large band of high cloud extending towards the pole from 40° S to 70° S over Antarctica between 10° W and 30° W. Figure 5b shows the outgoing longwave radiation (OLR) map for the day; the signature of the cloud deck is clearly visible, centred at (60° S, 20° W). The OLR is 50–80 W m⁻² (25–40%) lower than that of the air surrounding the cloud deck. Note the correspondence with the back trajectories in Fig. 4; cloud forms as air ascends and moves towards the pole. Some textural features over the ice cap suggest that the cloud deck may extend to 80° S, but the lack of contrast between the ice cap and the cloud leaves this as a tentative point. It is not obvious for this particular case whether the attenuation of OLR above the cloud deck is caused by the infrared opacity of the cloud, or by the coldness of its top. Experience^{4,16} with cloud decks over the Antarctic peninsula and Weddell Sea on 17 August and 5 September 1987, and over Lerwick and Scandinavia^{17,18} on 30–31 January 1989, suggests that these factors are correlated; a deep, strong flow of moist air towards the pole produces thick clouds with cold tops. From the point of view of the cloud's effect on the OLR flux, it is immaterial which of the effects is responsible.

The air-parcel trajectories show that the air near the South Pole in the upper troposphere on 24 July 1987 had originated at much lower latitudes within the previous five days, and had also ascended from higher tropospheric pressures. Thick clouds were formed, penetrating to polar latitudes. During the next 10 days, some of this air had returned to mid-latitudes and higher pressures. This mechanism of drying air by sloping convection² towards the pole during cyclogenesis in the circumpolar trough,

followed by return of the dried air to mid-latitudes as air slides down the Antarctic cold dome, accounts for the extra dryness of the austral mid-latitude upper troposphere in the winter, compared with its boreal counterpart. The temperatures in the cold air dome below 200 mbar over Antarctica are ~10 K colder than are found over the Arctic, and hence are capable of producing much drier air. Inspection of the isentropes on meridional cross-sections¹¹ shows that surfaces of constant potential temperature θ in the range 280 < θ < 310 K during austral winter connect the upper troposphere over Antarctica with the surface in the latitude belt from 55° S to the tropics. Sloping convection could therefore account for the extra dryness behind cold fronts in the Southern Hemisphere, which was reported recently by McMurdie and Katsaros¹⁹ from satellite microwave data on total column water vapour. Our observations are not readily comparable with those of Peixoto and Oort²⁰, because of the lack of precision and accuracy in the radiosonde data in the upper troposphere. They are consistent with the patterns of precipitation given by Jaeger²¹ in his Fig. 2b, which shows a belt of high precipitation near 60° S which has no boreal counterpart.

Ultimately, Antarctica is colder than the Arctic because it is a continent rather than an ocean, making it harder for ocean currents to transport heat to polar latitudes. Beyond this simple statement, however, lies a complex interplay between radiative cooling, sloping convection and the water content, both in clear air and cloud. Accurate simulation of these interactions is a severe test of general circulation models, even for the present atmosphere. Modelling the effect of a doubling of carbon dioxide on these factors, the sea surface temperatures, the pack ice and

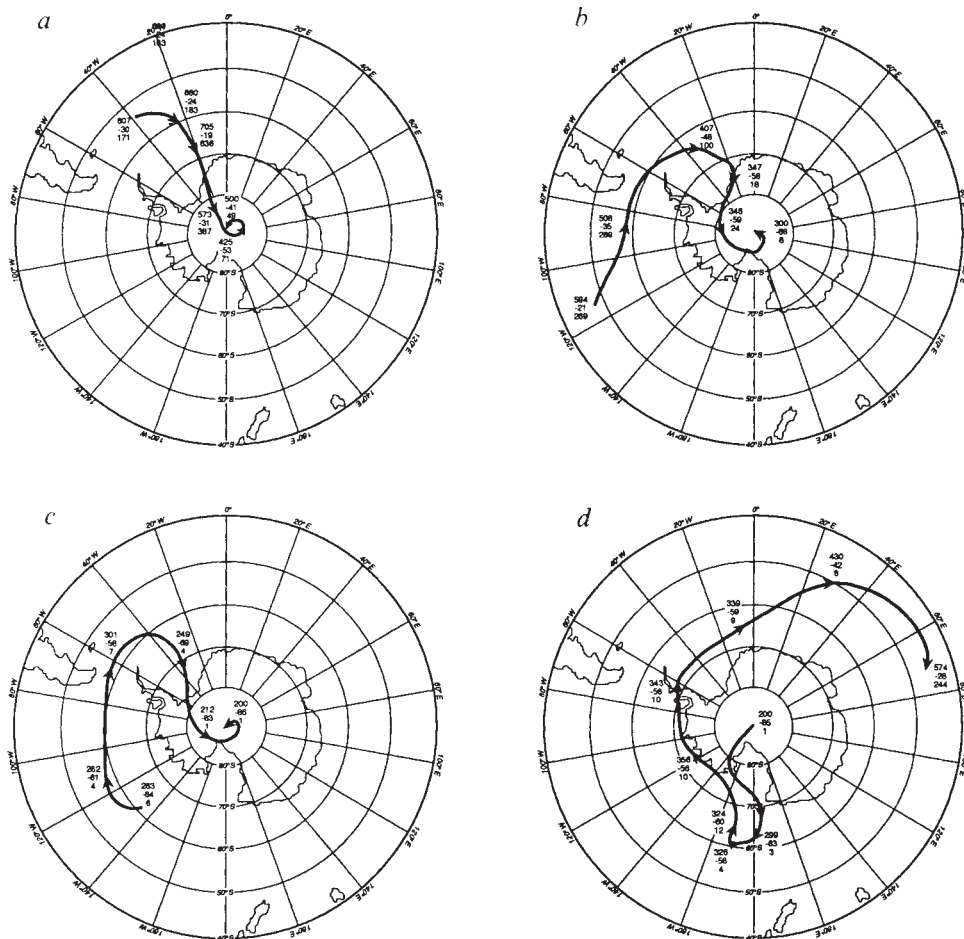
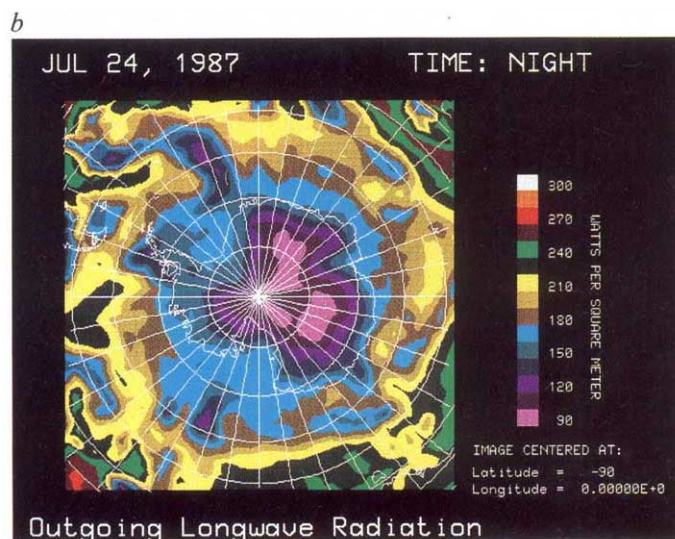
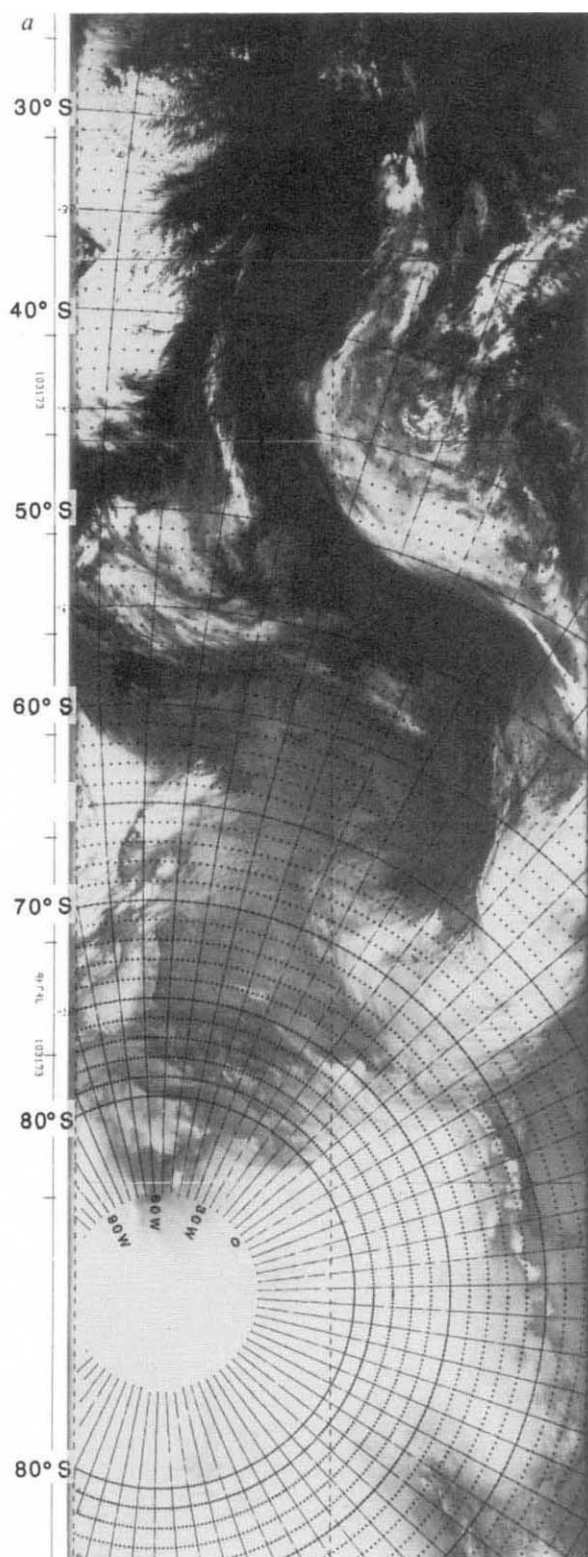


FIG. 4 Air-parcel trajectories ending (a–c) and beginning (d) at the South Pole on 24 July 1987 in the upper troposphere, calculated from ECMWF analyses. Arrows are every 24 h, and the numbers near each are pressure (mbar), temperature (°C) and humidity mixing ratio (p.p.m.v.). a, 500 mbar at Pole, back trajectory; b, 300 mbar at Pole, back trajectory; c, 200 mbar at Pole, back trajectory; d, 200 mbar at Pole, forward trajectory.



the wind fields will be even more demanding. It is clear, however, that our data need to be extended in coverage and to other seasons.

Intriguingly, the observed asymmetry in water vapour arises despite the symmetry in zonal mean sea surface temperatures in mid-latitudes at the times of measurement²², for example 5 °C at 50° latitude and 12 °C at 40° latitude. The mechanism of sloping convection² during cyclogenesis dries the poleward-moving air thereby providing a sink mechanism for water vapour; it is therefore crucial to the supply of water vapour in the upper troposphere, and hence to both cloud formation and the radiative balance. Such behaviour is consistent with the observation²³ that there is an excess of evaporation over precipitation at all latitudes except the sub-tropics. The subtropics are thus the source of water vapour for the troposphere; atmosphere dynamics convey moisture to middle and subpolar latitudes where it is precipitated. Antarctica is a more effective dry air factory than the Arctic, resulting in the observed interhemispheric wintertime asymmetry. It is clear that treatments of water vapour feedback under a CO₂ increase based solely on tropical sea surface temperature and convection are incomplete²⁴. □

Received 29 April; accepted 23 July 1991.

1. Tyndall, J. *Phil. Trans. R. Soc. A* **151**, 1–36 (1861).
2. Ludlam, F. H. *Clouds and Storms* Ch. 1 (Pennsylvania State University Press, 1980).
3. Doherty, G. M. & Newell, R. E. *Tellus* **B36**, 149–162 (1984).
4. Kelly, K. K. *et al. J. geophys. Res.* **94**, 11317–11357 (1989).
5. Vaghjiani, G. L. & Ravishankara, A. R. *Nature* **350**, 406–409 (1991).
6. Murphy, D. M., Kelly, K. K., Tuck, A. F., Proffitt, M. H. & Kinne, S. *Geophys. Res. Lett.* **17**, 353–356 (1990).
7. Kelly, K. K. *et al. Geophys. Res. Lett.* **17**, 465–468 (1990).
8. Somerville, R. C. J. & Remer, L. A. *J. geophys. Res.* **89**, 9668–9672 (1984).
9. Mitchell, J. F. B., Senior, C. A. & Ingram, W. J. *Nature* **341**, 132–134 (1989).
10. Komhyr, W. D. *et al. J. geophys. Res.* **94**, 11429–11436 (1989).
11. *Monthly Mean Aerological Cross-sections Pole to Pole Along Meridian 75° W for the IGY Period* (US Weather Bureau, 1961).
12. MacDowall, J. *Proc. R. Soc. A* **256**, 149–197 (1960).
13. Prantner, G. D. *World Met. Org. No.* **211**, 324–344 (1967).
14. Salter, P. R. S. & Merrick, S. D. *Met. Mag.* **118**, 59–63 (1989).
15. Schwerdtfeger, W. *Weather and Climate in the Antarctic* Ch. 4 (Elsevier, 1984).
16. Tuck, A. F. *J. geophys. Res.* **94**, 11687–11737 (1989).
17. Poole, L. R. *et al. Geophys. Res. Lett.* **17**, 389–392 (1990).
18. Jones, R. L., McKenna, D. S., Poole, L. R. & Solomon, S. *Geophys. Res. Lett.* **17**, 545–548 (1990).
19. McMurdie, L. A. & Katsaros, K. B. *Mon. Weath. Rev.* **119**, 589–605 (1991).
20. Peixoto, J. P. & Oort, A. H. *Variations in the Global Water Budget*, 5–66 (Reidel, 1983).
21. Jaeger, L. *Variations in the Global Water Budget*, 129–140 (Reidel, 1983).
22. Shea, D. J., Trenberth, K. E. & Reynolds, R. W. *NCAR tech. note TN-345 + STR* (1990).
23. Peixoto, J. P. *Nordic Hydrol.* **2**, 120–138 (1970).
24. Ramanathan, V. & Collins, W. *Nature* **351**, 27–32 (1991).

ACKNOWLEDGEMENTS. We thank the aircrew and ground staff of the NASA Ames Research Center, who operated the ER-2 and DC-8 aircraft during the AAO and AASE missions, and P. Salter, D. Merrick and B. Butler, the UK Meteorological Office forecasters at Punta Arenas and Stavanger.

FIG. 5 Satellite cloud data, Southern Hemisphere, 24 July 1987. *a*, Infrared image, F8 satellite, DMSF; note the correspondence of the cloud deck to the back trajectories in Fig. 4*a*, *b* and *c* between 10° W and 30° W. The latitude and longitude are in degrees south and degrees west. *b*, Outgoing longwave radiation; note the attenuation of the flux of OLR by the cloud deck stretching from (70° S, 20° W) to (40° S, 50° W). The data are from the archives at the National Snow and Ice Data Center, CIRES, University of Colorado, Boulder.

The photolysis of colloidal iron in the oceans

Mark L. Wells^{*†}, Lawrence M. Mayer[†],
Olivier F. X. Donard[‡], Marta M. de Souza Sierra[‡]
& Steve G. Ackelson[§]

[†] Department of Oceanography, University of Maine,
Darling Marine Center, Walpole, Maine 04573, USA

[‡] Laboratoire de Photophysique et Photochimie Moleculaire,
Universite de Bordeaux I, 33405 Talence Cedex, France

[§] Bigelow Laboratory for Ocean Sciences, McKown Point,
W. Boothbay Harbor, Maine 04575, USA

THE extent to which iron limits primary production in open ocean waters depends not only on the aeolian supply^{1–3}, but also on factors that control its availability for biological uptake. Although the marine chemistry of iron is poorly understood, much of it occurs in refractory particulate^{1,2} and colloidal⁴ states—forms unavailable for direct assimilation by phytoplankton^{5,6}. But iron availability depends on its chemical lability⁵, or ease of dissolution; hence processes that alter the lability of particulate and colloidal iron in sea water govern their availability to phytoplankton. Here we report that light increases the lability of colloidal iron in sea water of pH 8, with a photon-normalized spectral dependence that generally increases with decreasing wavelength from 400–300 nm. From optical modelling we predict that the incident solar spectrum, combined with the preferential attenuation of shorter ultraviolet wavelengths in sea water, will lead to a maximum depth-integrated photoreaction near 380–400 nm. Our results show that the photolysis of forms of solid iron may occur deep into the ocean's euphotic zone, and hence that the availability of iron to phytoplankton in the ocean may be much greater than previously thought.

Phytoplankton obtain iron by forming complexes of dissolved Fe(III) hydroxy species at membrane-bound uptake sites, followed by transport into the cell^{7,8}. Iron availability is therefore a function of the replenishment rate of Fe(III) hydroxy species to solution, in other words the dissolution kinetics, or chemical lability, of colloidal and particulate iron^{5,6}. The chemical lability of iron in sea water can be determined by reaction with the complexing agent 8-hydroxyquinoline (oxine); oxine-iron complexes formed in solution are extracted onto C₁₈ resin and the iron measured after elution by graphite furnace atomic absorption spectrometry⁵. There is a strong positive correlation between the fraction of iron complexed by oxine and the growth rates of marine algae in cultures, demonstrating that the reactivity of iron to form complexes with algal membrane transport ligands is covariant with its reactivity with chelators such as oxine⁵. Sunlight increases the oxine reactivities of both poorly crystalline ferrihydrites and the highly crystalline oxyhydroxide goethite in sea water⁹. (Goethite is one of the most stable of the iron oxyhydroxides and is widespread in terrestrial and marine systems.) This photoprocess apparently proceeds through a cycle of photoreductive dissolution/rapid reoxidation/precipitation involving organic chromophores, and yields amorphous, highly labile Fe(III) precipitates⁹ (Fe²⁺ is a transient species¹⁰ in oxic sea water of pH 8). But the oceanographic relevance of this photochemical cycling of iron largely depends on the spectral dependence of the photoreaction, which controls the penetration of iron photolysis into the euphotic zone.

We determined the photo-induced changes in iron lability at different wavelengths of ultraviolet and visible light by spiking sea water with colloidal iron of low lability and measuring its oxine reactivity after irradiation. We used surface sea water (5 m) from the Damariscotta estuary (Maine, USA), shelf water

(30 m) 30 km off the Gironde river estuary (France), and surface water (45 m) from the eastern margin of the Weddell Sea, Antarctica (75° 21.1' S, 25° 48.8' W). All seawater samples were collected using clean techniques that avoided contamination from the surface microlayer. Measurements of seawater absorbance and corrected¹¹ fluorescence spectra showed similar spec-

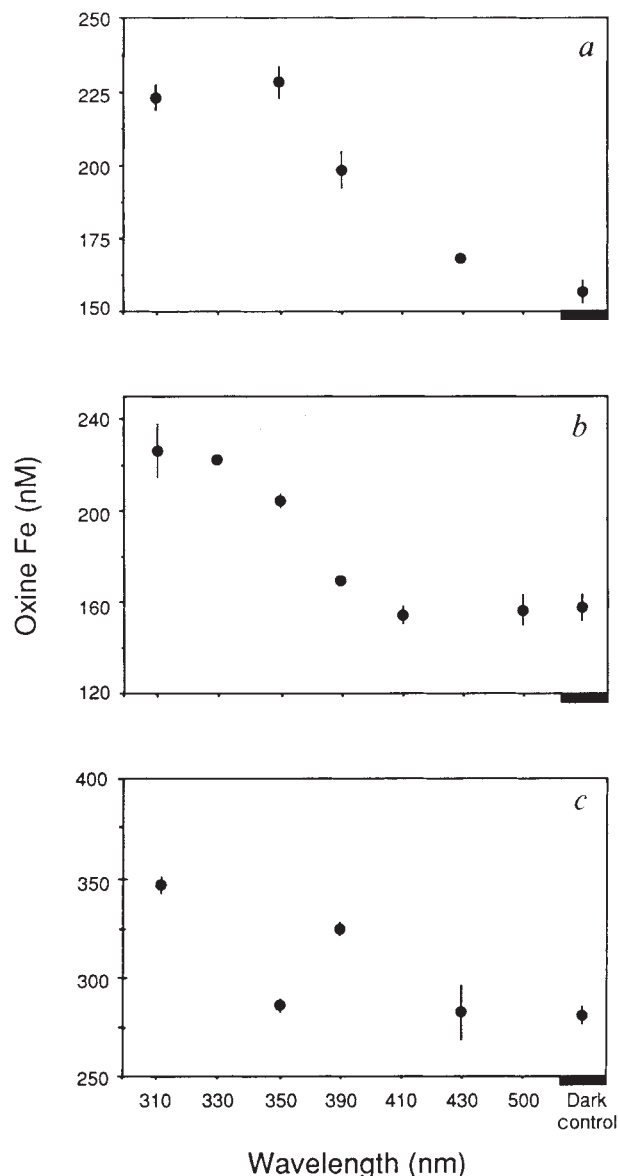


FIG. 1 The extent of photoreaction in 5 min of 90° FERR at different wavelengths in *a*, sea water from the Damariscotta river estuary, *b*, sea water off the Gironde River estuary and *c*, Antarctic sea water. Freshly prepared 90° FERR (see text) was added to the different sea waters (4 µM ferrihydrite) and irradiated in quartz-windowed, 10-cm spectrophotometer cells with an electric arc light source. Light was filtered through an adjustable monochromator to give 10 nm bandwidths centred at various ultraviolet and visible wavelengths. Solution temperatures were maintained at 9.5 ± 0.5 °C during irradiation. After irradiation, oxine was added to the cells and incubated for one hour at pH 6 (22 °C) before extracting the oxine-iron complexes from duplicate subsamples onto C₁₈ columns. The columns were then eluted with methanol and the iron determined by graphite furnace atomic absorption. Complete details of the methodology are given in Wells *et al.*⁵. The mean and range of duplicate analyses are shown. The detection limit was 50 nM iron with a precision of ±10%. The differences in the oxine reactivity of 90° FERR among the dark controls were probably due to minor temperature differences in iron stock before heating to 90 °C. Light intensities were measured at the specific irradiation wavelengths by chemical actinometry¹⁶; the changes in iron lability shown here are normalized to photon flux.

* Present address: Marine Research Division, A-020, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA.

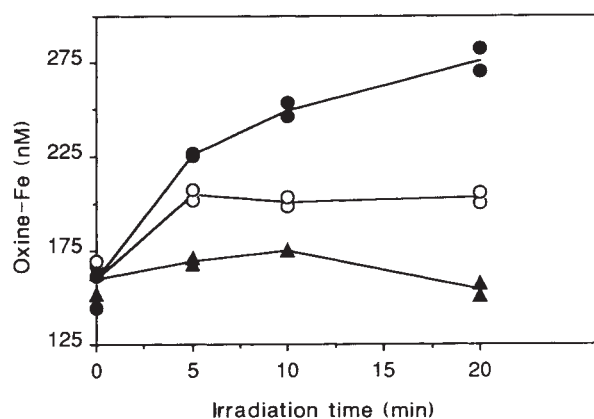


FIG. 2 Photoinduced changes in the oxine reactivity of 90° FERR in Gironde sea water. Freshly prepared 90° FERR (see text) was added to Gironde sea water (4 μ M ferrihydrite) and irradiated at 310 nm (●), 350 nm (○) and 390 nm (▲). The irradiation conditions were as described in Fig. 1. The changes in iron lability shown here are normalized to photon flux.

tral patterns but decreasing intensities (and therefore organic concentrations) in the order Damariscotta > Gironde > Antarctic. The iron substrate used in these experiments was synthetic colloidal ferrihydrite produced by heating freshly precipitated ferrihydrite at 90 °C for 5 min (90° FERR³). This ferrihydrite was selected because it is relatively unavailable to phytoplankton⁵, its chemical lability is similar to measured values of particulate iron lability in coastal and offshore sea waters¹², and its lability remains constant in the dark⁹.

Irradiation increased the oxine reactivity of 90° FERR in each of the three sea waters, the photolysis intensifying with decreasing ultraviolet wavelengths below 400 nm but absent at higher (photosynthetically active) wavelengths during the 5-min irradiation (Fig. 1). This wavelength dependence is very similar to that for the photoproduction of free radicals in surface sea water, which are short-lived intermediaries in marine photochemical transformations¹³. The iron photoreaction persisted longer at shorter wavelengths (Fig. 2). Pre-irradiating the sea-water with strong ultraviolet light eliminated the photoreaction

(not shown), suggesting that the reaction, and its wavelength dependence, was caused by organic chromophores. These chromophores probably form surface complexes on iron oxyhydroxides and cause ligand-to-metal charge transfers on photo-oxidation. Continuing photolysis presumably requires the sustained adsorption of new chromophores from solution. There were only slight changes in seawater absorbance (230–600 nm) or fluorescence spectra (emission scans of 320–600 nm with excitation at 313 and 370 nm) during the photoreaction experiments, and no clear losses of identifiable chromophores or fluorophores were detected as a result of the narrow-band irradiations that we used. The nature and direct source of the organic chromophores involved in iron photolysis are therefore unknown.

The photon-normalized reaction rates measured here are probably not representative of natural conditions; wavelength-specific photon fluxes were 10 to 100 times greater than for natural sunlight. This should cause faster photodegradation of the organic chromophores, and thus more rapid cessation of the photoreaction. In addition, the concentrations of 90° FERR added were two to three orders of magnitude higher than found in coastal and open ocean waters, respectively. The resultant, unrealistically high iron:chromophore ratios possibly lead to underestimation of the rate and extent of iron photolysis. More detailed studies of the photolysis kinetics of 90° FERR and colloidal goethite with simulated and natural sunlight are reported elsewhere⁹. In any event, the relative changes in 90° FERR lability over short (5-min) irradiation intervals (Fig. 1) demonstrate a spectral dependence for the photoreaction that is qualitatively similar among these diverse environments, and suggest that iron photolysis occurs commonly in surface waters.

The depth distribution of the relative rate of iron photolysis in sea water was modelled using the 5-min, photon-specific reaction rates of 90° FERR (Fig. 1) and the calculated, downwelling, diffuse irradiance in sea water. We modelled two water types having optical properties typical of coastal and open ocean waters; the sole difference between these being the extent of light attenuation due to organic matter. The resultant light fields were combined with the Damariscotta photolysis data in Fig. 1a. The Damariscotta data can be used to obtain a reasonable first approximation of colloidal iron photolysis in open ocean conditions because photolysis rate constants of 90° FERR in Damariscotta and Sargasso sea water under normal sunlight are nearly identical⁹. Our model does not account for the depth

Relative oxine-Fe production
(10^{-14} mol m⁻² s⁻¹)

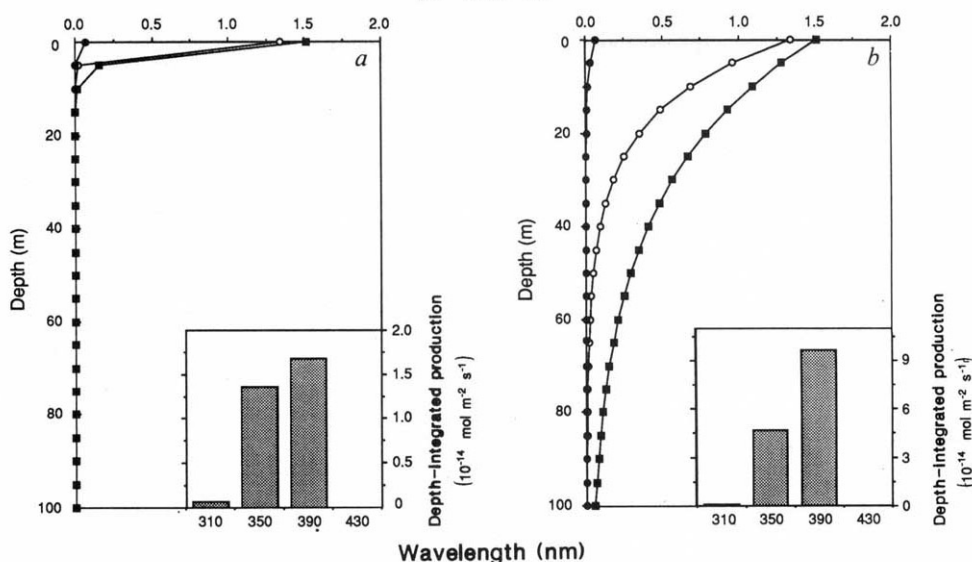


FIG. 3 The relative depth dependence of colloidal iron photolysis in a, coastal and b, open ocean sea water calculated from the photon-specific reaction rates and the estimated downwelling diffuse irradiance at 310 nm (●), 350 nm (○) and 390 nm (■). The optical model was constructed using mid-day, mid-latitude solar irradiances at sea level under a clear sky¹⁷. To determine the in-water diffuse attenuation coefficients at these wavelengths, we summed absorption coefficients for clear water¹⁸ and gelbstoff¹⁹ and light scattering due to clear water¹⁸. The relative oxine-iron photoproduction rates with depth were calculated as the product of the 5-min photon-specific photolysis rate of 90° FERR determined from Fig. 1a and the calculated vertical distribution of photon flux²⁰.

dependence of chromophore concentrations or any other depth or site-specific factor that might affect rate or extent of photo-reaction.

From the model results, as relative photoproduction rates of oxine-reactive iron (Fig. 3), we predict that colloidal iron photolysis would be concentrated in the upper 5–10 m in coastal waters, but would occur to depths >100 m in open ocean waters. In the open ocean, then, iron photolysis may operate over much of the depth range of photosynthetically active phytoplankton. Although these rate data must be considered to be preliminary (the light source and iron concentrations in these experiments were substantially different from natural conditions) the predicted depth distribution of iron photolysis closely matches the newly measured Fe(II) abundance in profiles of clear ocean water¹⁴. The depth-integrated production shows a reversal in the relative importance of wavelengths; substantially more colloidal iron photolysis would be associated with wavelengths near 390 nm than near 310 nm because of the lower incident spectral abundance and higher in-water attenuation of shorter ultraviolet wavelengths. Concentrations of chromophores that are activated by longer ultraviolet wavelengths (350–400 nm) are then likely to be the important cause of iron photolysis in sea water. Incubation experiments testing for iron limitation in sea water therefore should be conducted in containers that reproduce the ambient ultraviolet light conditions for the depth of interest.

Our experimental data and optical modelling, in conjunction with other experiments showing that highly crystalline Fe(III)

oxhydroxides such as goethite also are photolabile⁹, support the hypothesis that the photolysis of colloidal and particulate iron extends deep into the euphotic zone in open ocean environments. The availability of iron therefore may be much greater than estimated by the simple chemical dissolution of iron from particulates such as aeolian dust. □

Received 31 January; accepted 29 July 1991.

1. Martin, J. H. & Fitzwater, S. E. *Nature* **331**, 341–343 (1988).
2. Martin, J. H. & Gordon, R. M. *Deep Sea Res.* **35**, 177–196 (1988).
3. Martin, J. H., Gordon, R. M., Fitzwater, S. & Broenkow, W. W. *Deep Sea Res.* **36**, 649–680 (1989).
4. Mill, A. J. B. *Envir. Tech. Lett.* **1**, 97–108 (1980).
5. Wells, M. L., Mayer, L. M. & Guillard, R. R. L. *Mar. Chem.* **33**, 23–40 (1991).
6. Rich, H. W. & Morel, F. M. M. *Limnol. Oceanogr.* **35**, 652–662 (1990).
7. Anderson, M. A. & Morel, F. M. M. *Limnol. Oceanogr.* **27**, 789–813 (1982).
8. Morel, F. M. M. & Hudson, R. J. M. in *Chemical Processes in Lakes* (ed. Stumm, W.) 251–281 (Wiley, New York, 1985).
9. Wells, M. L. & Mayer, L. M. *Deep Sea Res.* (in the press).
10. Millero, F. J. & Sotolongo, S. *Geochim. cosmochim. Acta* **53**, 1867–1873 (1989).
11. Donard, O. F. X. *et al. Mar. Chem.* **27**, 117–136 (1989).
12. Wells, M. L. & Mayer, L. M. *Mar. Chem.* **32**, 195–210 (1991).
13. Zafiriou, O. C. & Dister, B. *J. geophys. Res.* **96**, 4939–4945 (1991).
14. O'Sullivan, D., Hanson, A., Miller, B. & Kester, D. *Limnol. Oceanogr.* (in the press).
15. Wang, M. K. & Hsu, P. H. *Soil Sci. Soc. Am. J.* **44**, 1089–1095 (1980).
16. Heller, H. G. & Langan, J. R. *J. chem. Soc. Perkin II*, 341–343 (1981).
17. Gast, P. R., Jursa, A. S., Castelli, J., Basu, S. & Aarons, J. *Handbook of Geophysics and Space Environments* (ed. Valley, S. L.) 16.1–16.38 (Air Force Cambridge Research Laboratory, Office of Aerospace Research, 1965).
18. Smith, R. C. & Baker, K. S. *Appl. Opt.* **20**, 177–184 (1981).
19. Bricaud, A., Morel, A. & Prieur, L. *Limnol. Oceanogr.* **26**, 43–53 (1981).
20. Kirk, J. T. O. *Geophys. Res.* **93**, 10897–10908 (1988).

ACKNOWLEDGEMENTS. We thank A. Monaco and R. Lesclaux for hospitality, M. Ewald for the Antarctic seawater sample, A. Dickson for helpful comments, and the CNRS ECOMARGE and EUROTRAC programs and the Center for Marine Studies, University of Maine, for funding this research.

An automaton for fractal patterns of fragmentation

Sandra J. Steacy & Charles G. Sammis

Department of Geological Sciences, University of Southern California, Los Angeles, California 90089–0740, USA

FRACTURES in the Earth's crust have a fractal structure over a wide range of length scales. A micromechanical model has been proposed¹ for the formation of fractal patterns of fragmentation in fault zones, based on the preferential fracture, at all length scales, of neighbours of a particle that have the same size as the particle itself. Here we explore this model in two and three dimensions using computer automata which implement these nearest-neighbour fracture rules. The automata produce random fractals which have capacity dimensions between 1.1 and 1.7 in two dimensions, and between 2.0 and 2.8 in three dimensions, the precise value depending on the packing geometry and the presence of long-range interactions imposed by uniform strain conditions. The fractal fragmentation patterns observed in natural systems tend to have dimensions between 2.5 and 2.7; we suggest that our model may permit an interpretation of these values in terms of the packing configuration (number of nearest neighbours) of the constituent particles.

Sammis *et al.*¹ showed that gouge from the Lopez Canyon fault zone in southern California was self-similar in the size range 10 μm to 1 cm with a fractal dimension $d_f = 1.6$ (in two-dimensional (2D) section, 2.6 for the corresponding three-dimensional (3D) isotropic random fractal). Barton and Hsieh² and Barton³ mapped 2-D fracture patterns on outcrops in the Nevada test site and found that they were fractal over scale ranges from centimetres to tens of metres with an average fractal dimension of 1.67 ± 0.07 .

Two processes that produce fractal fracture structures have been proposed, one related to the formation and growth of faults and the other to fragmentation in a previously granulated material. Both have been simulated in the laboratory. The growth

of shear fractures has been studied by Davy *et al.*⁴. They simulated the India–Asia continental collision by the slow advance of a rectangular indenter through a layer of cohesionless sand floating on a viscous substratum of silicone putty. The 2D pattern of conjugate shear fractures formed was found to be fractal with $d_f = 1.70 \pm 0.05$, independent of the viscosity of the substrate. Sornette *et al.*⁵ proposed that these patterns are the result of random crack growth in a critically loaded system, which is controlled by long-range screening and growth enhancement interactions between the growing fractures. In their model, entire domains of the crust are loaded to a critical state in an extension of the self-organized critical fault model of Bak and Tang⁶ to regional faulting patterns.

The development of a fractal fragmentation has been investigated by Biegel *et al.*⁷. They deformed a layer of initially uniformly sized fragments in simple shear and observed that the

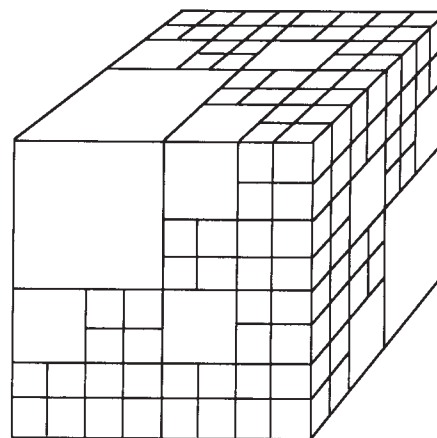


FIG. 1 Fractal cube. This figure has six blocks of edge length $1/2$ for every block of length 1, corresponding to a fractal dimension $d_f = \log 6 / \log 2 = 2.58$. The pattern has the property that no neighbours are the same size at any scale.

fragment sizes evolved to a fractal distribution with $d_f = 1.6$ in planar section. This observation was taken to support the proposal of Sammis *et al.*¹ that the 3D fractal dimension $d_f = 2.6$ is a direct consequence of the fracture probability of a particle being completely determined by the relative size of its nearest neighbours. They argued that a particle is most likely to fracture when it is loaded by other particles of the same size, as this geometry maximizes the stress concentrations on its surface. This hypothesis is supported by experiments in which Hoffman and Schönert⁸ loaded mixtures of glass beads in compression and found that the fracture probability of the larger spheres decreases as they are progressively surrounded by smaller spheres. The end result of a process that eliminates same-sized neighbours at all scales is a distribution in which no two particles the same size are nearest neighbours at any scale. As this is a property of the fractal pattern shown in Fig. 1, which has $d_f = 2.58$, Sammis *et al.*¹ suggested that this explained the fractal dimension observed in fault gouge.

Although the deterministic fractal in Fig. 1 has the requisite nearest-neighbour geometry (if neighbours are particles that share faces or edges), it is not obvious that a system that evolves according to a fracture probability based solely on the relative size of nearest neighbours will produce a random fractal with the same dimension. Here we explore the evolution of random fractals using computer automata which differ in the definition of neighbours and the presence or absence of strain constraints.

Each automata begins with an array of identical squares (in 2D) or cubes (in 3D) on which periodic boundary conditions are imposed. These blocks represent the uniformly sized particles with which Biegel *et al.*⁷ began their experiments. We use periodic boundary conditions to minimize boundary effects in a finite array. All three automata share the following rules. (1) A block's fracture probability is determined by the number of same-sized neighbours. Blocks with the largest number of such neighbours are most vulnerable to breakage. (2) Whenever two or more blocks have an equal probability of breakage, the choice of which one actually breaks is random. (3) In 2D, blocks break into four identically sized fragments, each 1/4 the original area. In 3D, cubes fracture into eight identical cubes. (4) Fragmenta-

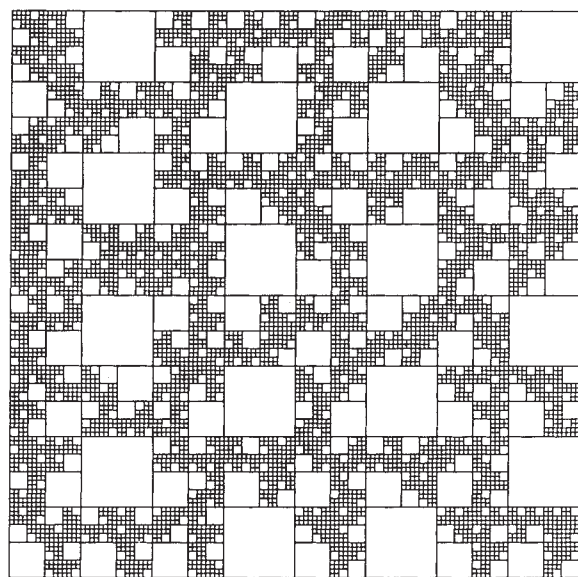


FIG. 2 Typical fractal pattern produced by the two-dimensional automata with eight possible neighbours and no stress bias. Automata began with an 8×8 array of identical squares, and fragmentation proceeded through five levels. The fractal dimension of this pattern is $d_f = 1.64$.

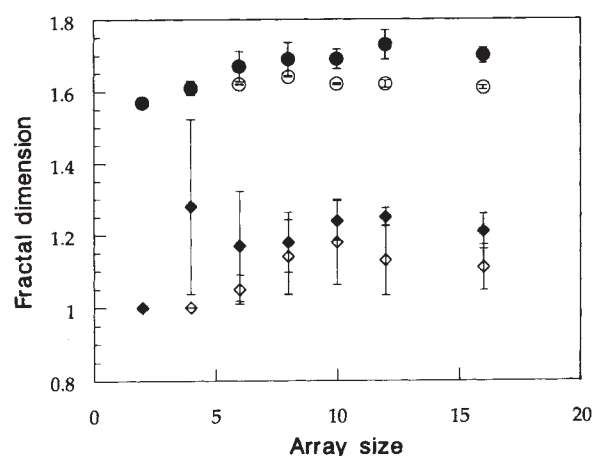


FIG. 3 Fractal dimension against array size for two-dimensional automata. ●, Eight neighbours; ○, eight, with bias; ◆, four neighbours; ◇, four, with bias. Each point represents the average of ten realizations, and the error bars show the standard deviation. As array sizes increase, the error bars decrease and d_f stabilizes. The fractal dimension is sensitive to the definition of neighbours and only weakly dependent on the stress bias.

tion begins with the largest blocks and proceeds to the next smaller scale only when none of the larger blocks are vulnerable to breakage.

We explore the effects of two variables on the resultant fragmentation: these are the packing configuration, as reflected in the definition of neighbours, and the presence or absence of long-range interactions which are imposed by a uniform strain constraint. This second constraint takes the form of a new rule in the automaton. Whenever two or more blocks are equally vulnerable (based on their neighbour configuration), the one that fractures is no longer determined by chance. Instead, the blocks that are close to a previously fractured block of the same size are more likely to fracture. If two or more blocks are still equally vulnerable, the choice is again made at random. The rationale for such a stress bias is that fracture of a block produces a stress concentration in its immediate neighbourhood which biases the fracture probabilities of the remaining blocks. The effect is to distribute the deformation more uniformly over the array.

Neighbours are defined in terms of the geometry of a square or cubic array. A cubic lattice was chosen because squares or cubes can be fractured into smaller squares or cubes while still filling space. In 2D, we consider two configurations. (1) Neighbours are blocks sharing either edges or corners; there are eight maximum possible neighbours for any given particle. (2) Neighbours are blocks sharing edges only; there are a maximum of four. We consider three configurations in 3D. (1) Neighbours share faces, edges or corners; there are 26 maximum. (2) Neighbours share faces or edges; there are 18 maximum (this is the geometry of the perfect fractal in Fig. 1). (3) Only blocks sharing faces are neighbours; the maximum is 6. The cubic array therefore only allows us to examine a limited number of nearest neighbours (6, 18 or 26). If a space-filling scheme based on another geometric shape could be devised, it would be possible to investigate other numbers of nearest neighbours. We conjecture that such results would be consistent with the trends established below for the cubic array, but this remains unproven. The number of nearest neighbours should be regarded in a more general sense. Clearly, the presence of 26 identically sized neighbours is specific to cubic particles and is much larger than likely values for more irregularly shaped particles. But nearest neighbours may be interpreted as the number of same-sized blocks

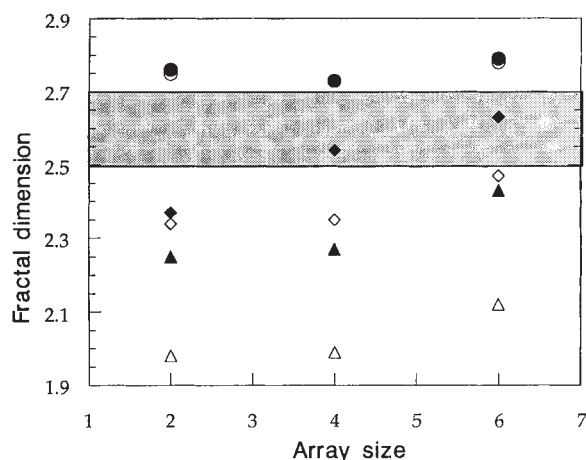


FIG. 4 Fractal dimension against array size for three-dimensional automata. ●, 26 neighbours; ○, 26, with bias; ◆, 18 neighbours; ◇, 18, with bias; ▲, 6 neighbours; △, 6, with bias. As in Fig. 3, each point represents the average of ten realizations; the error bars are similar to those for two dimensions and have been omitted for clarity. The fractal dimension is sensitive to the definition of neighbours and, for all configurations except for 26 neighbours, to the stress bias. The stippled region indicates the range 2.6 ± 0.1 observed in natural fault gouge.

seen by any given block during deformation, and this need not be limited by grain shape or instantaneous packing geometry.

Figure 2 shows a typical realization of the 2D automaton in which particles sharing edges or corners were considered to be neighbours, and no stress bias was imposed. This pattern is fractal with dimension $d_f = 1.64$, similar to that of the gouge produced by Biegel *et al.*⁷

The results of the 2D automata are summarized in Fig. 3, in which d_f is plotted as a function of the size of the starting array. Each point is the average of ten realizations, and the error bars indicate the standard deviation. In general, as the array size increases, the standard deviation decreases and d_f stabilizes. The fractal dimension is sensitive to the definition of nearest neighbours and only weakly dependent on the stress bias. Note that the fractal dimension we are measuring is the capacity dimension, which depends solely on the particle size distribution. A measure of the correlation dimension⁹ would probably be more sensitive to the stress bias, as this uniform strain condition affects the spatial distribution of deformation.

The results of the 3D automata are summarized in Fig. 4. Again, each point is the average of ten realizations. The error bars are similar to those in Fig. 3 and have been omitted for clarity. Note that the fractal dimension increases with the size of the starting array. The same effect is observed in 2D, for which it is possible to run large enough arrays to achieve stability. Although this has not been possible in 3D, comparison with the 2D case suggests that the 6×6 3D array may be close to equilibrium. As in 2D, D_f is sensitive to the definition of neighbours, but it is also sensitive to the stress bias for all cases except when neighbours are defined as sharing faces, edges and corners. The stippled band in Fig. 4 indicates the range 2.6 ± 0.1 observed in natural fault gouge¹. It appears that packing configurations that have 26 neighbours produce a larger d_f than observed, whereas configurations with ~ 18 neighbours can produce random fractals with the observed dimension.

From these automata results, we draw the following general conclusions. A fragmentation process that eliminates nearest neighbours at all scales can produce a random fractal distribution. The fractal dimension depends on the packing geometry and on long-range interactions imposed by a constraint of uniform strain. Those models (3D) that have ~ 18 neighbours can

simulate observed fragmentations. Although the long-range order imposed by a uniform strain condition affects the fractal dimension, it is not required to simulate the observed capacity dimension. □

Received 4 March; accepted 23 July 1991.

1. Sammis, C. G., King, G. & Biegel, R. *Pure appl. Geophys.* **125**, 777–812 (1987).
2. Barton, C. C. & Hsieh, P. A. *Am. geophys. Union Guidebk* T385, (1989).
3. Barton, C. C. in *Fractals and their Use in the Earth Sciences* (eds Barton, C. C. & LaPointe, P. R.) (GSA Memoir, in the press).
4. Davy, P., Sornette, A. & Sornette, D. *Nature* **348**, 56–58 (1990).
5. Sornette, A., Davy, P. & Sornette, D. *Phys. Rev. Lett.* **65**, 2266–2269 (1990).
6. Bak, P. & Tang, C. *J. geophys. Res.* **94**, 15635–38 (1989).
7. Biegel, R. L., Sammis, C. G. & Dieterich, J. H. *J. struct. Geol.* **11**, 827–846 (1989).
8. Hoffman, N. & Schöner, K. *Aubereit Tech.* **12**, 513–518 (1971).
9. Baker, G. L. & Gollub, J. P. *Chaotic Dynamics* 112–119 (Cambridge University Press, 1990).

Relationship of deep seismicity to the thermal structure of subducted lithosphere

George Helffrich* & John Brodhoft

Department of Geology, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK

RECENT experimental work on silicate olivine polymorphs¹ has confirmed earlier observations of mechanical failure in analogous compounds^{2,3} as a result of phase transformations when non-hydrostatic stresses were applied to a metastable phase. This 'transformational faulting'⁴ mechanism is considered to be a leading candidate⁵, among others involving olivine transformation^{6,7}, for the cause of deep earthquakes. Evidence consistent with this mechanism comes from the observation^{2–4} that, worldwide, earthquakes become more numerous with increasing depth after a seismicity minimum at about 350 km depth⁸. This depth lies within the range anticipated for mineralogical transformations in the subducted lithosphere⁹. But individual subduction zones differ in their thermal structures, so if transformational faulting indeed contributes to deep seismicity, this should be reflected in the location of the seismicity minimum for each zone. The depth at which the minimum occurs should depend on temperature in the same way as do the polymorphic transformations of olivine, and should always lie deeper than the depth at which the equilibrium transformation takes place. Here we test these predictions for eight subduction zones worldwide. We find that, with one exception (the North Japan zone), the depth of the seismicity minimum decreases linearly with increasing thermal age of the slab (a measure of its temperature profile). These results support the proposal that deep earthquakes are a consequence of phase transformations in olivine.

Deep earthquakes are enigmatic, since materials are believed to be too strong to undergo brittle fracture at great depth¹⁰, yet they apparently rupture along fault planes as occurs in shallow earthquakes, where the displacements may be observed directly. Many different explanations for the increase in deep seismic activity in subduction zones have been advanced. These include the interaction of the slab with the higher-viscosity lower mantle¹¹, the concentration of the slab-pull stresses into an elastic core that is narrowing with depth as the slab warms^{12–14}, amorphization reactions¹⁵ and the grain-size-related material strengthening that follows a phase transformation whose fine-grained reaction products initially weaken the material and

* Present address: Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington, DC 20015 USA.

permit aseismic deformation by superplastic flow⁶. An emerging alternative view (reviewed by Kirby *et al.*⁴) associates this increase in seismicity with the mechanism of 'transformational faulting' which is only possible at depths beyond the equilibrium olivine-spinel transition. In this model, material failure occurs during a kinetically delayed phase transformation when the nuclei of the stable phase coalesce to an extent related to the ambient stress state. We explore the consequences of a model of olivine phase change by relating the deep seismicity increase to the pressure and temperature conditions in individual subduction zones. The worldwide seismicity data set that has been cited as supporting this model is dominated by events from a single subduction zone, Tonga. If the seismicity increase in individual subduction zones are related to phase changes, they

should depend on the pressure and temperature conditions particular to each.

Two extreme temperature structures of subducting lithosphere can be envisaged that encompass all actual subduction zones. Young ocean floor near mantle temperature will maintain this temperature during subduction. We would expect a seismicity increase after the olivine phase transformation is reached at a depth of ~ 400 km, the depth that is generally ascribed to the olivine $\alpha \rightarrow \beta$ transformation in the mantle^{9,16}. At the other extreme, the earliest phase transformation in olivine within old, thoroughly cooled ocean floor subducted rapidly with insignificant heating will occur at a depth of ~ 200 km, as the Clapeyron slopes for reactions between the olivine α , β and γ polymorphs are all positive⁹. A useful parameter to characterize

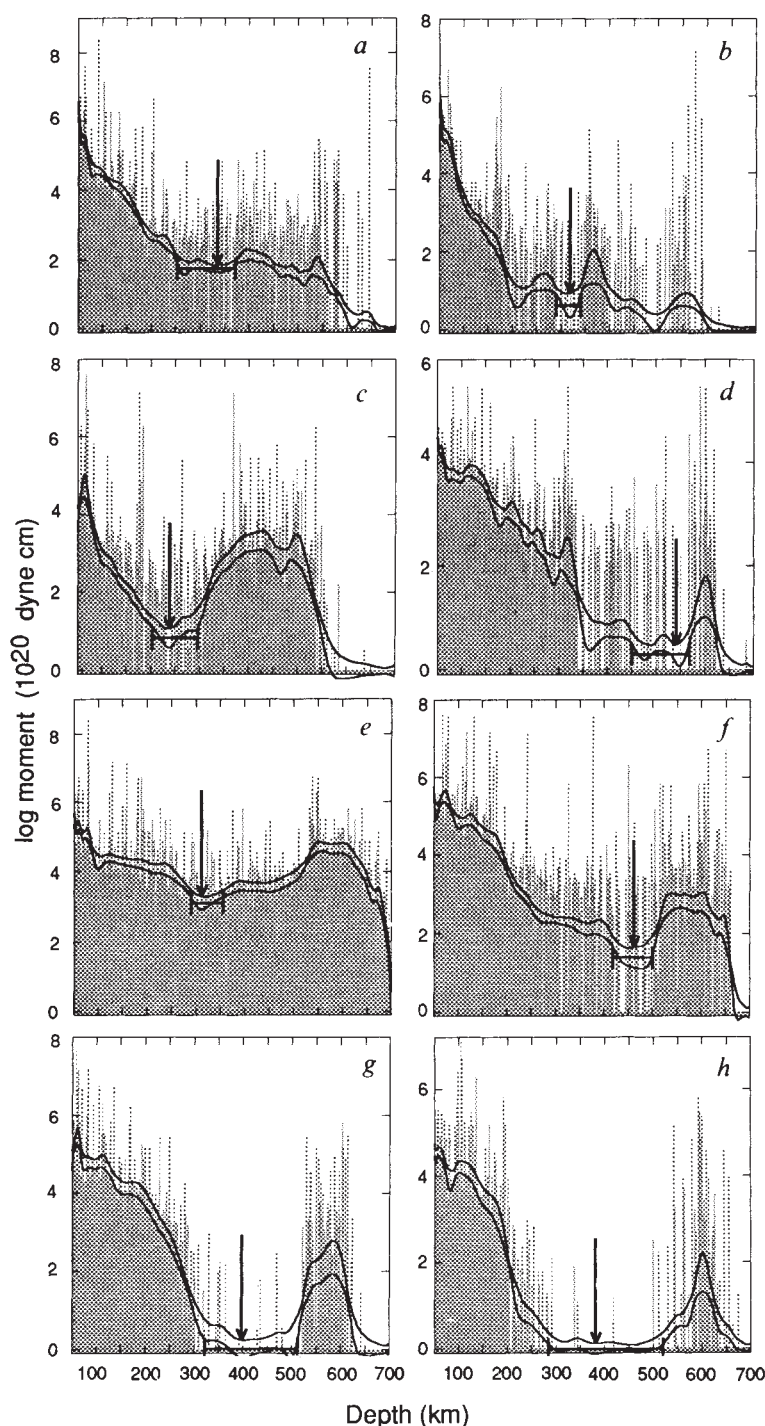


FIG. 1 1960–1989 moment release (dyne cm) within each subduction zone as a function of depth. Moment calculated from body-wave magnitude m_b (ref. 23) and Chebyshev-smoothed²⁴ (solid lines are mean $\pm$ variance of orders $n = 4$ –45). The first local minimum demanded by the variance bounds defines the seismicity minimum (marked with arrow), and the persistence of the minimum moment value within the variance bounds defines the bracket. Only seismicity deeper than 50 km is shown, to avoid dominating the profiles with shallow seismicity. a, Kuril; b, North Japan; c, Izu-Bonin; d, Marianas; e, Tonga; f, Java; g, Peru; h, Chile.

TABLE 1 Data used to model depth profiles of seismicity

Region*	Dip† (°)	Convergence rate‡ (km Myr ⁻¹)					Age§ (Myr)			Depth of deep seismicity increase (km)		
		Average	Spot		Rate	Azimuthal angle (°)	Low	High	Pref.	Low	High	Pref.
			Lat. °N	Lon. °E								
Kuril	45	82.5	48	157	82.6	112	84	119	100	253	371	335
North Japan	30	95.7	37	143	96.3	108	94	140	125	291	341	318
Izu-Bonin	51	57.9	30	143	58.1	107	110	160	140	204	297	238
Marianas	59	36.7	17	147	41.5	128	134	170	170	450	568	541
Tonga	47	63.3	-25	185	71.7	87	84	120	100	291	356	312
Java	56	74.2	-9	110	76.6	21	57.8	144	119	418	500	462
Peru	34	76.1	-10	282	77.9	80	23	66.4	48	285	520	382
Chile	35	83.4	-25	289	83.9	78	36.6	66.4	58	322	512	397

* Names from Jarrard²⁰, but Kuril includes Kamchatka, Java includes Sumatra and Andaman, and Peru includes Ecuador and Colombia.

† Dip is an average dip defined by the arc sine of the depth of the deepest seismicity within the region divided by the down-dip slab length. The down-dip lengths for each slab were obtained from the projection of a swath of the seismicity between 300 and 1,500 km width onto a trench-normal transect fixed at the point at which the spot convergence rates were computed.

‡ Average rate is the trench-normal NUVEL-1 convergence rate²² integrated along the plate boundary where subduction is proceeding, and divided by the arc length of the boundary normal to the convergence vector. This value is used in the thermal age calculation. Spot rates given for comparison.

§ Age range from range of seafloor chrons¹⁸ at trench. Where age of deepest seismically active portion of the slab is younger²⁰, this age is taken as a bound. 'Pref.' indicates preferred value.

these aspects of the thermal structure of the subducted lithosphere is the thermal age¹⁷, the product of convergence rate and lithosphere age. By this measure, the greater the thermal age, the cooler the slab. We expect from the above examples an inverse dependence of seismicity minimum on the thermal age, with a shallower seismicity minimum in thermally old subduction zones and a deeper minimum in thermally young ones.

To obtain depth profiles of seismicity we used the NGDC data tape (comprising the US Geological Survey Preliminary Determination of Epicenters catalogue and its ancestors). We examined the seismicity of subduction zones having seismicity deeper than 300 km and consuming material of known age. Age bounds for the subducted lithosphere were estimated from seafloor ages^{18,19} and from ages of subducted material²⁰. The convergence rates used to calculate thermal ages are along-strike averages. Table 1 contains the raw data.

In each subduction zone we obtain the depth at which seismicity increases using smoothed moment release profiles (Fig. 1). We isolate the persistent features in them using the variance bounds from point averages of successively smoother profiles. Our points are chosen at the shallowest points where the variance bounds require a local minimum. Data from depth profiles of earthquake counts yield similar results, so we believe

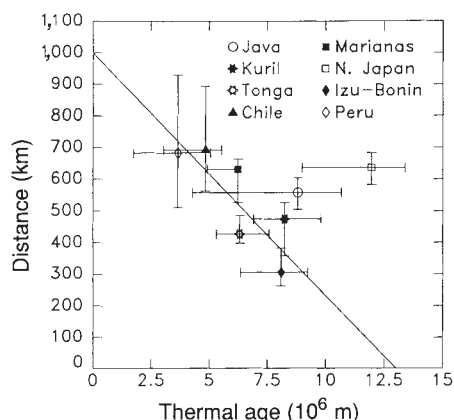


FIG. 2 Down-dip distance over which subducted material has been transported to the depth of increasing seismicity, shown as a function of its thermal age, age $\times$ rate. Line is consistent with all error brackets, except for Northern Japan. Uncertainties include depth uncertainty (Fig. 1) and thermal age uncertainty (Table 1).

that the conclusions do not depend on the methodology used. The distance travelled by the subducted material to the depth of the seismicity minimum is shown as a function of thermal age in Fig. 2. There is a negative correlation, suggesting that whatever the cause of the increase in the numbers of deep earthquakes, it has a thermal dependence similar to that of the phase transformations in olivine. The value for North Japan falls significantly off the trend, perhaps as an effect of the decreasing age of the subducted slab with depth¹⁸ (K. Petronotis, personal communication) or because it is the thermally oldest (compare with other thermal age-based regressions^{17,20}).

We can make a more direct test that phase transformations among the olivine polymorphs control the increase in deep seismicity by checking whether the increase always occurs below the equilibrium transformation depth in the slab. As all of these phase-transition pressures decrease with decreasing temperature^{9,16}, the minimum slab temperature gives the shallowest transformation depth.

To obtain slab temperatures, we use a finite-difference temperature model incorporating mantle flow^{10,21}, with dips, ages

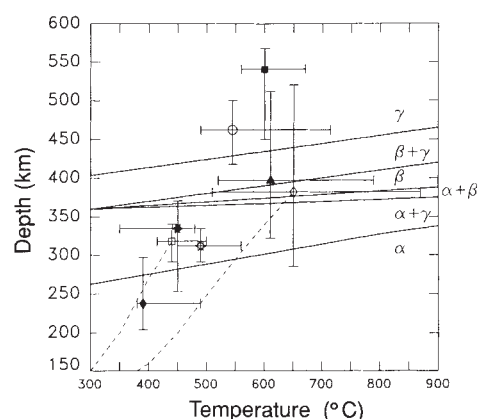


FIG. 3 Minimum temperature within each subduction zone as a function of depth; symbols as in Fig. 2. Full lines indicate the equilibrium phase boundaries from Akaogi *et al.*⁹ involving α , β and γ in $(\text{Mg}_{0.89}\text{Fe}_{0.11})_2\text{SiO}_4$ and extrapolated to low temperatures. Temperatures calculated with a finite-difference temperature model¹⁰. No latent heat of phase change is included in the temperature calculations. Dashed lines show minimum temperatures in Northern Japan and Peru slabs to indicate pressure-temperature trajectories; paths in other slabs lie between these two.

and convergence rates from Table 1. We show the minimum slab temperatures at the depth of the increase of deep seismicity in each subduction zone in Figure 3. With the possible exception of Izu-Bonin, all deep seismicity increases occur beyond the $\alpha \rightarrow \alpha + \gamma$ transition within the slab, again consistent with the hypothesis that olivine phase transformations mediate deep seismicity. The overall pattern is one of two equally sized groups, one at the lower left lying close to the $\alpha \rightarrow \alpha + \gamma$ phase boundary and the other farther away. The former group's proximity to the phase boundary suggests that in some cases seismicity increases closely follow phase transformations. Whether these are due to the changes in the slab stress state caused by the phase transformation or due to an unchanged stress state in a new, weaker slab rheology is unclear. Whatever the mechanism, the idea that olivine phase transformations somehow control deep earthquakes is supported by the seismicity patterns in individual subduction zones. It is not, however, obvious how this occurs. $\square$

Received 17 April; accepted 15 July 1991.

- Green, H. W., Young, T. E., Walker, D., Scholz, C. & Prior, D. *Trans. Am. Geophys. Un.* **71**, 946 (1990).
- Kirby, S. H. *J. geophys. Res.* **92**, 13789–13800 (1987).
- Green, H. W. & Burnley, P. C. *Nature* **341**, 733–737 (1989).
- Kirby, S. H., Durham, W. B. & Stern, L. A. *Science* **252**, 216–225 (1991).
- Kirby, S. H. *Trans. Am. Geophys. Un.* **71**, 946 (1990).
- Rubie, D. C. *Nature* **308**, 505–508 (1984).
- Lomnitz-Adler, J. *J. Phys. Earth* **38**, 83–98 (1990).
- Frohlich, C. *Ann. Rev. Earth planet. Sci.* **17**, 227–254 (1989).
- Akaogi, M., Ito, E. & Navrotsky, A. *J. geophys. Res.* **94**, 15671–15685 (1989).
- Brodholt, J. & Stein, S. *Geophys. Res. Lett.* **15**, 1081–1084 (1988).
- Vassiliou, M. S., Hager, B. H. & Raefsky, A. *J. Geodynam.* **1**, 11–28 (1984).
- Wortel, R. *Nature* **296**, 553–556 (1982).
- Wortel, R. *Geophys. Res. Lett.* **13**, 34–37 (1986).
- Wortel, R. & Vlaar, N. J. *Pure appl. Geophys.* **128**, 625–659 (1988).
- Meade, C. & Jeanloz, R. *Science* **252**, 68–72 (1991).
- Bina, C. R. & Wood, B. J. *J. geophys. Res.* **92**, 4853–4866 (1987).
- Molnar, P., Freedman, C. & Shih, J. F. *Geophys. J. R. astr. Soc.* **56**, 41–54 (1979).
- Larson, R. L. et al. *The Bedrock Geology of the World (map)* (Freeman, New York, 1985).
- Leg 129 shipboard scientific party *Nature* **345**, 112 (1990).
- Jarrard, R. D. *Rev. Geophys.* **24**, 217–284 (1986).
- Toksöz, M. N., Sleep, N. H. & Smith, A. T. *Geophys. J. R. astr. Soc.* **35**, 285–310 (1973).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. *Geophys. J. Int.* **101**, 425–478 (1990).
- Geller, R. J. *Bull. seism. Soc. Am.* **66**, 1501–1523 (1976).
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes* (Cambridge University Press, 1987).

ACKNOWLEDGEMENTS. We benefited from discussions with J. Holloway, C. Jaupart, and B. Wood, and thank S. Kirby for a preprint. The comments of the reviewers of an earlier manuscript, S. Kirby and S. Stein, improved this one considerably.

Spatial structure and chaos in insect population dynamics

Michael P. Hassell*, Hugh N. Comins*
& Robert M. May†

* Department of Biology and Centre for Population Biology,
Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY, UK

† Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

MOST environments are spatially subdivided, or patchy, and there has been much interest in the relationship between the dynamics of populations at the local and regional (metapopulation) scales¹. Here we study mathematical models for host–parasitoid interactions, where in each generation specified fractions (μ_N and μ_P , respectively) of the host and parasitoid subpopulations in each patch move to adjacent patches; in most previous work, the movement is not localized but is to any other patch². These simple and biologically sensible models with limited diffusive dispersal exhibit a remarkable range of dynamic behaviour: the density of the host and parasitoid subpopulations in a two-dimensional array of patches may exhibit complex patterns of spiral waves or spatially chaotic variation, they may show static ‘crystal lattice’ patterns, or they may become extinct. This range of behaviour is obtained

with the local dynamics being deterministically unstable, with a constant host reproductive rate and no density dependence in the movement patterns. The dynamics depend on the host reproductive rate, and on the values of the parameters μ_N and μ_P . The results are relatively insensitive to the details of the interactions; we get essentially the same results from the mathematically-explicit Nicholson–Bailey model of host–parasitoid interactions, and from a very general ‘cellular automaton’ model in which only qualitative rules are specified. We conclude that local movement in a patchy environment can help otherwise unstable host and parasitoid populations to persist together, but that the deterministically generated spatial patterns in population density can be exceedingly complex (and sometimes indistinguishable from random environmental fluctuations).

Insect parasitoids lay their eggs on, in or near the bodies of other arthropods, and the parasitoid larvae kill the host as they feed on it. The dynamics of systems with discrete but synchronized generations of hosts and parasitoids are thus simpler than most other prey–predator associations (see equation (1)) and make convenient systems for study³. Theoretical and empirical studies of predator–prey systems in general have contributed much to the emerging consensus that spatial patchiness is important to the understanding of regulation and persistence of many natural populations^{4–11}. Recent work on the dynamics of host–parasitoid interactions in particular has also begun to provide a unifying framework for understanding the magnitude of the patch-to-patch variability in levels of parasitism that is required for the population as a whole to persist^{12–14}. Little attention, however, has been given to the importance of the type of dispersal between units of local population density (but see refs 10, 11). This work is devoted largely to this question, and to the surprising answers that can emerge when dispersing individuals only move locally (rather than spreading globally, as in most previous work).

In a homogeneous environment, host–parasitoid dynamics may be described by a pair of first-order difference equations

$$N_{t+1} = \lambda N_t f(N_t, P_t) \quad (1a)$$

$$P_{t+1} = q N_t (1 - f(N_t, P_t)). \quad (1b)$$

Here N_t and P_t are, respectively, the host and parasitoid population sizes in generation t ; λ is the average number of

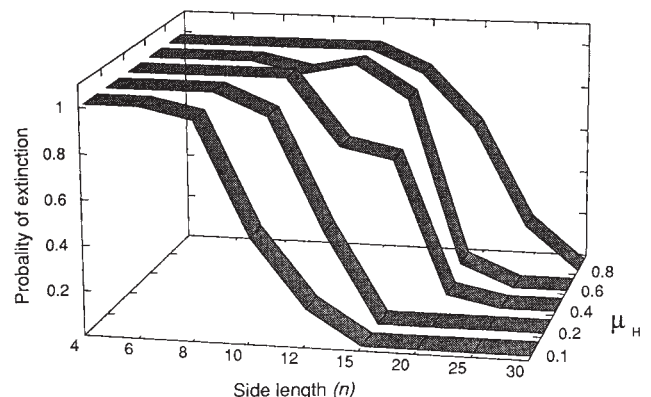


FIG. 1 Extinction probabilities for the specific model described in the text, in relation to the numbers of patches in a square grid of side length (n) and the fractions of hosts dispersing to neighbouring patches (μ_N) ($\mu_P = 0.89$, $\lambda = 2$, reflective boundaries). Extinction is measured as the proportion of 20 replicates failing to persist over 2,000 generations. Each replicate is started by setting nonzero population densities in only the third patch from the left in the top row. The same 20 pairs of initial host and parasitoid densities are used for all the parameter combinations. Local extinction occurs by numeric underflow (densities less than about 10^{-45}); however, the results are robust when local extinction thresholds for both hosts and parasitoids are modelled explicitly.

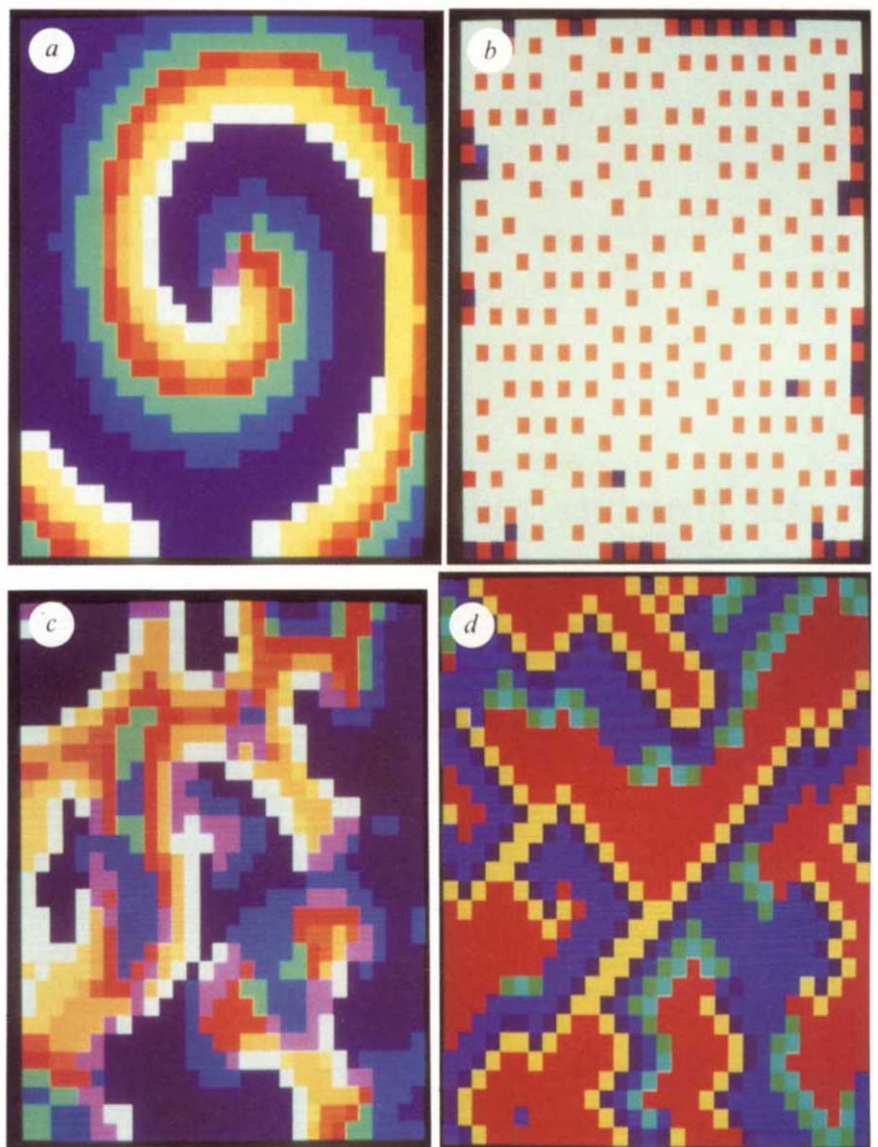
offspring produced by an unparasitized host (assumed to be independent of host density; stability considerations will be more complicated when λ has significant density-dependence); $f(N, P)$ is the fraction of hosts that escape parasitism; and q is the average number of female parasitoids emerging from each host parasitized. If parasitoids search randomly and independently, then we get the classic Nicholson–Bailey model ($f = \exp(-aP)$), which is unstable with diverging oscillations¹⁵. If we suppose that the environment consists of a square grid containing many patches, then the dynamics involves two steps. In the first step, we apply equation (1) to the subpopulations in each patch, to find how many hosts and parasitoids will emerge in that patch in the next generation (the ‘escape function’ f in any one patch may be of simple Nicholson–Bailey form, corresponding to random search within patches, or it may be more complex). In the second step, we allow for dispersal, distributing some fixed fraction (μ_N) of the hosts and (μ_P) of the parasitoids in each patch among other patches, according to some specified rules. In most previous studies, the dispersing individuals are distributed according to some particular statistical distribution (often a negative binomial) over all other patches. In these cases, overall population densities will tend to be stable provided the coefficient of variation squared of the density of searching parasitoids in the vicinity of each host exceeds approximately unity (the ‘ $CV^2 > 1$ rule’)^{12–14}. This

approximate criterion only depends on the degree of heterogeneity in the risk of parasitism between individual hosts¹⁶; thus spatial patterns of parasitism that are directly or inversely related to host abundance per patch, or that show no such covariance, can all contribute to population persistence in the same way¹⁴. When we look at the two-dimensional spatial array of densities of host and parasitoid subpopulations generated by such ‘global movement’ models, we simply see the spatially stochastic pattern corresponding to the underlying statistical distribution used to describe the dispersal process.

Suppose, however, rather than the dispersing individuals from each patch entering a ‘pool’ for global dispersal, they diffuse outwards, in this case by equal distribution among the eight nearest neighbours of that patch in the square array. To lay bare this effect alone, moreover, let us suppose parasitism is random in each patch ($f = \exp(-aP)$ in equation (1)). For specified values of the parameters λ , μ_N and μ_P (the parameters q and a are principally scaling factors and otherwise do not affect the dynamics), we now have a purely deterministic model for the spatial and temporal dynamics of this host–parasitoid association. We have conducted extensive numerical and analytical studies of this system in which the size of the $n \times n$ array of patches is varied. The results are as follows.

If the number of patches is too small, then the underlying instability of the simple Nicholson–Bailey model in each patch

FIG. 2 Photographs illustrating the different patterns of spatial dynamics obtained from the specific model (a–c) and the cellular automaton model (d) discussed in the text. Each photograph is a snapshot in time with the colour coding representing different relative abundances of hosts and parasitoids within a patch. a, Typical ‘spiral waves’ obtained in the spiral region of Fig. 3. b, The ‘crystal lattice’ pattern obtained for $\mu_P \rightarrow 1$ and small μ_N (top left of Fig. 3). This pattern settles to a completely static mosaic of high density and low density patches; there is variation within the high and low density categories, although this variation does not show up with our colour-coding. c, Spatially erratic (‘chaotic’) patterns obtained in the chaos region of Fig. 3. d, A typical spatial pattern generated from the cellular automaton in which the movement rules correspond qualitatively to the specific model. The automaton has nine states, labelled A to I; movement to the next state in cyclic order is automatic, except that state A (empty) moves to state B only in the presence of at least one neighbouring B (modelling host colonization), and state D moves to state E only in the presence of an F neighbour (modelling parasitoid colonization). Only the four orthogonal nearest neighbour cells are used. Variant automata (exhibiting spirals, crystals and other behaviour) are generated by using eight nearest neighbours, and by changing the required neighbours for $A \rightarrow B$ to C or D and for $D \rightarrow E$ to D, F, G or H. These changes affect the velocities of the colonization wavefronts of hosts and parasitoids, and are analogous to changes in μ_N and μ_P in the specific model.



eventually drives the entire system to extinction; there are diverging oscillations in local densities, and first host and then parasitoid populations are extinguished. What is meant by 'too small' an array depends on the fraction of hosts and parasitoids that leave their home patch each generation (μ_N and μ_P , respectively), and on the host's intrinsic growth rate, λ . Figure 1 illustrates this, showing the proportion of simulations that result in extinction as a function of μ_N and of the side-length, n , of the array of patches. For instance, with $\mu_N = 0.1$ all the arrays of 6×6 or smaller result in extinction, all of 15×15 or greater persist, whereas for intermediate-sized arrays there is an element of chance (with some initial configurations leading to persistence and others to extinction, sometimes after quite long times). Higher degrees of host movement (higher μ_N) make persistence less easy and, as ever, persistence is also more difficult for very large λ . These results are all fairly insensitive to whether one assumes periodic, reflexive, or absorbing boundary conditions on the array (the details of boundary conditions will be described elsewhere (H.N.C., M.P.H. and R.M.M., manuscript in preparation)).

The persisting deterministic systems settle to one of three broadly-distinguishable patterns.

The first shows patterns of spiral waves. Figure 2a shows a snapshot in time of such a pattern (assuming reflection at the boundaries). Over time, the waves spiral outwards, collide, and perform all manner of manoeuvres, but all the time retaining their overall wavelength and the spiral pattern. Previous studies of the spatial dynamics of prey-predator interactions in continuous time (as distinct from our discrete-time interactions with difference equations) have employed nonlinear 'reaction-diffusion' differential equations, and have found stable wave patterns (see, for example, ref. 9 for a study of goldenrod aphids and ladybug beetles, and ref. 17 for a more general review). The complex spiral waves in Fig. 2a are, however, new in this context.

The second pattern, attained for a relatively restricted range of values of μ_P near 1 and small μ_N , gives a completely static 'crystal lattice', as illustrated in Fig. 2b. Analytical results support

the finding that these 'crystal lattice' patterns need not be neatly periodic in space; they can be deformed, spatially irregular lattices, while retaining their essential property of being temporally static (the static 'crystal lattices' are, moreover, only possible if λ is small enough—specifically, $\lambda < e = 2.7$).

The third pattern, illustrated by Fig. 2c, is one of spatial chaos, with the spatially erratic pattern changing from generation to generation in an apparently unpredictable way. The underlying equations, however, are rigidly deterministic, with no random or unpredictable elements whatsoever (in contrast to the spatial patterns of the 'global dispersal' models^{18,19}, in which the distribution of the dispersing individuals among patches is stochastic, described by some statistical distribution). Figure 2c is an example of spatial chaos, generated by a simple and biologically motivated model that is wholly deterministic; the temporal sequence of successive patterns is even more striking.

Figure 3 indicates roughly which of the three characteristic patterns we may expect to see for particular values of μ_N and μ_P , assuming a 30×30 array of patches. This figure is necessarily impressionistic because the 'crystal lattice' pattern of Fig. 2a gradually shades into the spatial chaos of Fig. 2c, which in turn gradually shades into the spiral wave pattern of Fig. 2b as μ_N increases and μ_P decreases. Finally, no matter how large the two-dimensional array of patches, the system goes extinct if $\mu_N \rightarrow 0$ or $\mu_P \rightarrow 0$.

Turning back to the overall population dynamics, these results for models with local movement may be related to the previously discussed results based on global dispersal. Obviously, the results support the general point that patchiness, coupled with local movement among patches, can stabilize host-parasitoid associations (note that the underlying interaction in any one patch is by itself highly oscillatorily unstable). A. M. De Roos, E. McCauley and W. G. Wilson have independently made this point in an analysis of the overall dynamics of host-parasitoid systems similar to those studied here (personal communication). In more detail, we find that the rough stability rule mentioned above, $CV^2 > 1$, is supported in a qualified way by the models in our study: the associations which persist by virtue of spatial chaos or spirals (and which have roughly steady overall population values) consistently have CV^2 slightly greater than one. On the other hand, static crystal lattices have $CV^2 \ll 1$.

How robust are these results? To what extent do the patterns in Fig. 2, and their parameter-dependence in Fig. 3, depend on the detailed assumption of Nicholson-Bailey dynamics (with random search) within each patch? We could attempt to answer these questions by exploring other specific forms of the function $f(N, P)$ in equation (1). This is the procedure that has usually been adopted in past studies of this general kind³. Instead, we adopt the more novel approach of using a cellular automaton to explore the robustness of our conclusions about the population dynamics²⁰⁻²². In the work described above, we used an explicit equation for the dynamics in each patch and explicit formulae describing movement to neighbouring patches. To display the results, the two-dimensional arrays of population densities were colour-coded in Fig. 2 by assigning nine colours to characterize nine categories of patches defined by dividing each population density into three classes ('very low, medium, high'). For a cellular automaton approach, we abandon detailed host and parasitoid population values, and acknowledge only these nine qualitative categories of patch densities. We then define a set of 'movement rules' that specify the colour of each patch next generation in relation to its present colour and the colour of its eight neighbours. These rules are qualitative embodiments of the workings of dynamics within a patch and subsequent dispersal patterns.

Thus defined, our cellular automata give results qualitatively very similar to those of the detailed models. Arrays that are too small lead to extinction. Large enough arrays can give crystal lattices, or spiral waves, or spatial chaos (in which deterministic rules, with no random elements, give apparently random spatial

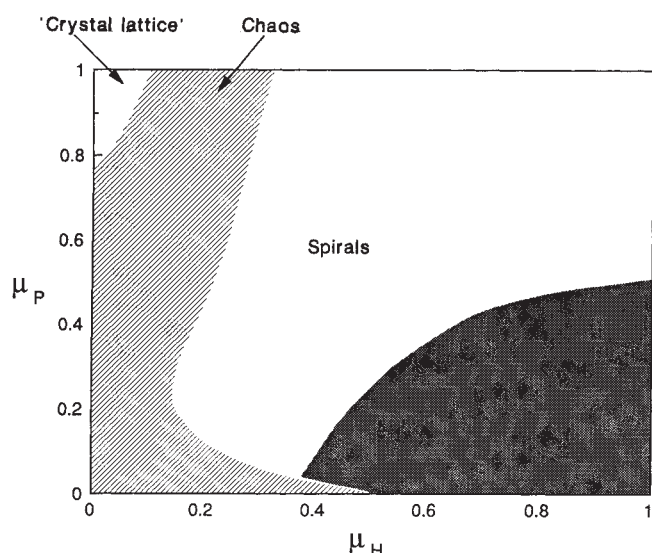


FIG. 3 Diagram showing the dependence of the type of persistent spatial pattern observed on μ_N and μ_P , for $n=30$ and $\lambda=2$. The boundaries are obtained by simulation, and are approximate (and partly subjective). The shaded area represents parameter combinations for which the persistent spatial pattern is unlikely to be established by starting the simulation with a single non-empty patch (as described in Fig. 1 caption). Spirals may be established in these cases by starting with a lower μ_N and increasing it after 100 to 200 generations. Non-persistence occurs for some combinations with very small μ_N or μ_P ; this area is imperceptible in the figure.

patterns). Figure 2d is a representative example, being the cellular automaton analogue of Fig. 2a with spiral waves moving as 'fronts' across the patchy environment.

In a homogeneous environment, host-parasitoid associations with discrete generations will not persist unless the parasitoids search nonrandomly, or unless host or parasitoid populations exhibit some form of inherent density-dependent effects. In a subdivided environment, such associations can persist without any explicit density-dependences, and with parasitoids searching randomly within patches, provided the hosts in each generation are sufficiently aggregated in their distribution among patches and there is sufficient variability in the density of searching parasitoids in the vicinity of individual hosts. We have extended these results to show that purely deterministic processes, whereby some constant fractions of hosts and parasitoids (μ_N and μ_P , respectively) move to immediately neighbouring patches in each generation, can also lead to overall persistence, even when parasitoids search randomly within patches and no explicit density-dependent mechanisms are present. The densities of host and parasitoid subpopulations in a two-dimensional ($n \times n$) array of patches can exhibit spiral waves, or spatial chaos (with deterministic rules generating patterns that are apparently random in time and space), or static-looking 'crystal lattice' patterns, depending on the magnitude of the parameters, μ_N and μ_P , and provided n is big enough. Note that these complex patterns of spatial variation in population density arise even though the environment in our different patches is the same; the patterning is thus intrinsically generated by the interplay of local dispersal and local dynamics. These results seem robust, being true both for conventional mathemati-

cal models based on Nicholson-Bailey or similar equations, and for cellular automata based on qualitative rules. We emphasize that these complex patterns in the densities of host and parasitoid subpopulations in individual patches (as shown in Fig. 2c) are produced by strictly deterministic rules or equations. In practice, such spatially chaotic patterns may be hard to distinguish from those produced by randomly varying environmental factors. □

Received 20 May; accepted 7 August 1991.

1. Gilpin, M. E. & Hanski, I. *Metapopulation Dynamics: Empirical and Theoretical Investigations* (Academic, London, 1991).
2. Taylor, A. D. *Ann. Zool. Fenn.* **25**, 63-74 (1988).
3. Hassell, M. P. *The Dynamics of Arthropod Predator-Prey Systems* (Princeton University Press, Princeton, 1978).
4. Hassell, M. P. & May, R. M. *J. Anim. Ecol.* **42**, 693-726 (1973).
5. Hassell, M. P. & May, R. M. *J. Anim. Ecol.* **43**, 567-594 (1974).
6. Maynard Smith, J. *Models in Ecology* (Cambridge University Press, Cambridge, 1974).
7. Kareiva, P. *Lect. Not. Biomath.* **54**, 368-389 (1984).
8. Kareiva, P. *Nature* **321**, 388-391 (1989).
9. Kareiva, P. & Odell, G. M. *Am. Nat.* **130**, 233-270 (1987).
10. Nachman, G. *J. Anim. Ecol.* **56**, 247-265 (1987).
11. Nachman, G. *J. Anim. Ecol.* **56**, 267-281 (1987).
12. Pacala, S., Hassell, M. P. & May, R. M. *Nature* **344**, 150-153 (1990).
13. Hassell, M. P. & Pacala, S. *Phil. Trans. R. Soc.* **330**, 203-220 (1990).
14. Hassell, M. P., Pacala, S., May, R. M. & Chesson, P. L. *Am. Nat.* **138** (in the press).
15. Nicholson, A. J. & Bailey, V. A. *Proc. Zool. Soc. London* **3**, 551-598 (1935).
16. Chesson, P. L. & Murdoch, W. W. *Am. Nat.* **127**, 696-715 (1986).
17. Murray, J. D. *Mathematical Biology* (Springer, London, 1989).
18. Hassell, M. P. & May, R. M. *Ann. Zool. Fenn.* **25**, 55-61 (1988).
19. Reeve, J. D. *Am. Nat.* **132**, 810-836 (1988).
20. Conway, J. H. *On Numbers and Games* (Academic, London, 1976).
21. Wolfram, S. *Nature* **311**, 419-424 (1984).
22. Crawley, M. J. & May, R. M. *J. theor. Biol.* **125**, 475-489 (1987).

ACKNOWLEDGEMENTS. This work was supported by the NERC Centre for Population Biology at Imperial College (H.N.C.), the Natural Environment Research Council (M.P.H.) and The Royal Society (R.M.M.).

A selective deficit for writing vowels in acquired dysgraphia

Roberto Cubelli

Rehabilitation Department, Ospedale Maggiore, Bologna I-40133, Italy

BRAIN-DAMAGED patients with acquired writing disorders provide important information about the normal processes of spelling and writing^{1,2}. Current models indicate that to produce a letter string, its 'abstract' representation is computed and stored in a temporary orthographic buffer, from which it is converted to a verbal code (if the word is to be spelled aloud) or to a physical letter code (if the word is to be written). The stored graphemic representations specify the identity and order of the component letters³ and their consonant/vowel status⁴. Here I describe the spelling performance of two patients with a selective deficit in writing vowels. When writing words, the first patient omitted all vowels, leaving a blank space between consonants or consonant clusters, whereas the second produced errors that almost exclusively involved vowels. This pattern of performance supports the hypothesis that the consonant/vowel status of graphemes is differentially specified in the spelling process and may be selectively affected after brain damage.

C.F. is a 43-year-old right-handed engineer with eight years of schooling. On 21 September 1990 he suffered an ischaemic infarction which involved the parietal lobe of the left cerebral hemisphere. On the first neuropsychological examination, two weeks after the stroke, he was totally speechless but could communicate by gestures. Auditory verbal comprehension was clinically normal. He wrote with the left hand because of a severe right hemiplegia. When asked to write his name and the names of his town and of five objects, he omitted all vowels (Fig. 1), leaving a blank space between correctly written consonants. Consonant clusters were preserved and no space was left between letters. C.F. was aware of errors, but did not seem

to be able to choose the correct letters. The deficit was transient. In the following days, the patient began to improve and a week later he showed a mild Broca's aphasia with articulatory difficulties and some anomic pauses. Writing performance was almost normal with only sporadic letter substitutions. Patients who leave blank spaces for unavailable letters are not uncommon^{5,6}, but a specific vowel impairment is completely novel.

C.W. is a 62-year-old right-handed man. He is a retired typographer with eight years of schooling. On 31 August 1990, he suffered an ischaemic infarction in the left frontal subcortical region. He had neither motor nor sensory deficits. C.W. showed a form of transcortical motor aphasia with some difficulties in initiating speech but relatively good auditory-verbal comprehension. Spontaneous speech was severely reduced with perseveration and occasional phonemic and verbal paraphasias, but no articulatory disorders. C.W. showed a specific deficit in writing (Table 1).

His spelling abilities were widely investigated (Table 2). He showed the same level of performance and the same kind of errors independently of input (oral dictation, delayed copying) or output (written spelling, oral spelling, typing) modalities. Writing was not affected by lexical factors (grammatical class, word frequency or abstract quality) or lexicality (words versus non-words). Only stimulus length influenced his performance, with short stimuli being spelled better than longer ones. The number of letters and not the number of syllables seemed to be the critical factor. C.W. made significantly fewer errors in writing capital letters than cursive. This is probably because with capital letters the patient was slower and made more corrections. In any case, the pattern of errors was not different in the two conditions. Direct copying and writing to dictation of single letters and syllables were flawless. C.W. wrote with his (dominant) right hand.

In the entire corpus of responses, 409 letters were produced incorrectly. The most frequent error type was letter substitution ($n=340$ (83%); for example, dietro (behind) → diatro). Other errors were transpositions ($n=30$ (7%); for example, caro (dear) → cora), deletions ($n=26$ (6%); for example, premio

TABLE 1 C.W.'s correct responses in reading aloud, writing to dictation and repetition of 40 words and 40 non-words (List 1)

	Words	Non-words
Writing	18/40 (45%)	24/40 (60%)
Reading aloud	38/40 (95%)	35/40 (87.5%)
Repetition	33/40 (82.5%)	34/40 (85%)

Words were controlled for their frequency, and how concrete they were⁹. Non-words were derived from words by substituting a single letter. A difference in performance was found ($\chi^2=36.878$; $P=0.0001$). Writing to dictation was significantly less accurate than reading aloud ($F=15.668$; $P<0.05$) or repetition ($F=10.19$; $P<0.05$). No difference was found between repetition and reading aloud ($F=0.587$; not significant).

(prize) → premo) and insertions ($n=13$ (3%); for example, uomo (man) → uoumo).

Substitution and transposition errors respected the consonant/vowel status of the graphemes involved in these errors. C.W. made 340 substitutions, in 336 cases (99%), vowels were substituted for vowels and consonants for consonants. Of the 30 transposition errors, 27 (90%) involved exchanging vowels for vowels and consonants for consonants.

C.W.'s deficit seems to be graphemic and not phonological in nature: he made almost no phonological errors in spontaneous speech and in oral naming; graphemic and not phonological units influenced his spelling performance: for example, he made errors like 'castagna (chestnut) → castanga', muschio (musk) → mushio', and 'plagio (plagiarism) → pliago'; even if the letter sequences (gn), (ch) and (gi) in Italian orthography correspond to single phonemes (respectively, /p/, /k/ and /dʒ/), they do not act as indissoluble units but are susceptible to errors like any other letter clusters (for example, 'borsa (bag) → bosra' and 'esiste (it exists) → esite').

The pattern of C.W.'s performance is indicative of a specific deficit of the graphemic buffer, namely of the structure occupying a central position between the processes that compute graphemic representations of words and nonwords, and peripheral output mechanisms.

It is known that the graphemic buffer can be selectively damaged^{3,4,6-8}, but to my knowledge a patient with a specific deficit in writing vowels has not been reported before. C.W. made significantly more errors with vowels than with consonants across all tasks (Table 3). No difference was found in writing stressed and unstressed vowels. The letter (u) was misspelled both when the corresponding phoneme was a vowel (for example, 'buio (dark) → boio') and when it was a consonant (vuoto (empty) → vioto'). The difference was not significant ($\chi^2=0.718$; not significant).

Caramazza and Miceli⁴ have also reported a patient who like C.W. systematically substituted vowels for vowels and con-

TABLE 2 C.W.'s performance on writing tasks

Stimuli	Correct responses	χ^2	P
List 1			
Words	90/140	1.497	NS
Non-words	80/140		
Short items (4 letters)	83/120	7.683	<0.01
Long items (7 letters)	62/120		
Concrete words	43/70	0.785	NS
Abstract words	48/70		
High-frequency words	46/70	0.031	NS
Low-frequency words	45/70		
Written spelling (cursive)	43/80		
Oral spelling	25/40	2.599	NS
Delayed copying	52/80		
Typing	51/80		
List 2			
Nouns	30/40		
Adjectives	26/40	1.505	NS
Verbs	30/40		
Functors	30/40		
Cursive	46/80	18.056	<0.001
Capital letters	70/80		
List 3			
Bisyllabic words	59/80	0	NS
Trisyllabic words	59/80		
Cursive	45/80	25.311	<0.001
Capital letters	73/80		

Only short stimuli from list 1 were given for oral spelling. List 2 included 20 nouns, 20 adjectives, 20 verbs and 20 functors. List 3 comprised 80 six-letter words: half stimuli were trisyllabic CVCVCV words, where C represents a consonant and V a vowel; half were bisyllabic complex CV words (CVVCVCV; CVCCVCV; CCVCVCV; CCVCVCV). All words contained 3 vowels and 3 consonants. Examples of C.W.'s errors: davanti (in front of) → devonta; perduto (lost) → pardeta; marsupio (marsupium) → morsupuo; auto (car) → uuto; dama (checkers) → de... du... da... doma; vilonta (non-word) → velento; invedia (non-word) → invadei; pianura (plain) → pianore; nuovo (new) → nouva; giusto (right) → gianste; finisce (it ends) → finisce. NS, Not significant.

TABLE 3 C.W.'s writing errors according to letter types

Stimuli	Vowels	Consonants	χ^2	P
List 1				
Written spelling	42	16	17.578	<0.001
Oral spelling	18	3	11.906	<0.001
Delayed copying	35	7	24.543	<0.001
Typing	25	15	5.221	0.02
Lists 2, 3, 4 and 5				
Cursive	147	22	109.179	<0.001
Capital letters	48	4	40.461	<0.001
Written naming	24	3	22.569	<0.001

List 4 comprised 60 words, 30 with a vowel cluster (for example, piuma) and 30 containing a vowel sequence with a 'non-phonological' grapheme (for example, the (i) of 'giunta'). There was no difference in frequency of errors like 'gioco (play) → geuco' and 'pendio (slant) → pendeo', despite the fact that the grapheme (i) has a phonological content only in the latter case. List 5 included 56 trisyllabic words, half stressed on the second syllable (regular words) and half stressed on the first syllable (irregular words). There was no difference in accuracy in writing stressed vowels (99/112) and unstressed vowels (202/224) ($\chi^2=0.255$; not significant).

sonants for consonants; he also made transposition errors that involved exchanging consonants for consonants and vowels for vowels. On this basis, they proposed that graphemic representations have a multidimensional structure, and that the consonant/vowel status of a grapheme is part of the graphemic representation of a word. For example, the graphemic representation of the word 'testa' would be {(t-C), (e-V), (s-C), (t-C), (a-V)}. When the identity of a grapheme is not specified, its

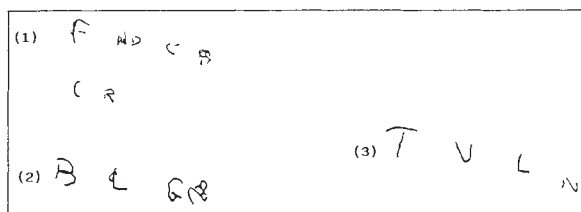


FIG. 1 Examples of C.F.'s written performances. (1) FONDACARO CIRO (patient's name) → FND C R C R; (2) BOLOGNA → B L GN; (3) TAVOLINO (little table) → T V L N.

consonant/vowel information may affect the selection of graphemes with the same consonant/vowel status. In the patients described here, the identity of vowels seems to be selectively affected, for example {<t-C>, <_V>, <s-C>, <t-C>, <_V>}. The preservation of the vowel specification in C.F. produced a total inability to choose the target letter ('t s t'), whereas in C.W. it led to incorrect choices, that is, to letter substitutions ('toste').

The specific deficit in writing vowels shown by C.F. and C.W. suggests that the identity of vowel and consonant letters is differentially computed by the brain and may be selectively affected by a cerebral lesion. This conclusion implies that in the orthographic system as well, the consonant/vowel opposition is not just a formal distinction but reflects a psychological reality. □

Received 20 May; accepted 24 July 1991.

1. Ellis, A. W. *Normality and Pathology in Cognitive Functions* (Academic, London, 1982).

2. Margolin, D. J. Q. *J. exp. Psychol.* **316A**, 459-489 (1984).

3. Caramazza, A. et al. *Cognition* **26**, 59-85 (1987).

4. Caramazza, A. & Miceli, G. *Cognition* **37**, 243-297 (1990).

5. Morton, J. in *Cognitive Processes in Spelling* (ed. Frith, U.) 117-133 (Academic, London, 1980).

6. Katz, R. B. *Cortex* **27**, 111-119 (1991).

7. Posteraro, L. et al. *Brain Lang.* **35**, 274-286 (1988).

8. Hillis, A. E. & Caramazza, A. *Brain Lang.* **36**, 208-235 (1989).

9. Sartori, G. *La Lettura* (Il Mulino, Bologna, 1984).

ACKNOWLEDGEMENTS. I thank P. Nichelli and A. Caramazza for their comments on earlier versions of this paper.

Identification and mapping to chromosome 1 of a susceptibility locus for periinsulinitis in non-obese diabetic mice

Henri-Jean Garchon, Pierre Bedossa*, Laure Eloy & Jean-François Bach

INSERM U 25, Hôpital Necker, 161 Rue de Sévres 75743 Paris Cedex 15, France
*Service d'Anatomie Pathologique, Hôpital Antoine Béchère, Clamart 92140, France

INSULIN-DEPENDENT diabetes mellitus (IDDM) is a polygenic disease caused by autoimmune destruction of insulin-producing β cells in the islets of Langerhans^{1,2}. Its onset is preceded by a long and variable period in which lymphoid cells infiltrate the pancreas but first remain outside the islets (peri-insulinitis) before invading them (insulinitis)³⁻⁷. Among susceptibility loci, only the major histocompatibility complex (MHC) has been clearly assigned⁸⁻¹². Genetic study of the nonobese diabetic (NOD) mouse model for insulin-dependent diabetes mellitus has revealed genetic linkage of insulinitis and of early onset diabetes with two non-MHC loci mapping to chromosome 3 and 11 respectively¹³. Here we report a close association of periinsulinitis with a third non-MHC locus mapping to chromosome 1. Successive stages in the progression of diabetic disease thus appear to be controlled by distinct genes or sets of genes.

In addition to the pancreas, certain exocrine glands in NOD mice such as salivary, lacrimal, peribronchial and vaginal glands, are consistently affected by cellular infiltration (refs 14, 15 and P.B., unpublished data). These inflammatory infiltrates consist of focal accumulation of lymphoid cells around vessels and secretory ducts and thus closely resemble periinsulinitis. This is very similar to Sjögren's syndrome in humans¹⁶. All these glands share common physiological and anatomical features as well as antigenic determinants with the exocrine pancreas^{17,18}. Added to the simultaneous cell-mediated transfer of sialitis and insulinitis in NOD neonates¹⁹, these observations strongly suggest that at

TABLE 1 Linked inheritance of periinsulinitis and sialitis in (NOD × B6)F₂ and (NOD × NZW)F₂ mice

Pathological trait		(NOD × B6)F ₂ cross			(NOD × NZW)F ₂ cross		
Peri-insulinitis	Sialitis	Female	Male	Total	Female	Male	Total
+	+	8	9	17	43	22	65
−	+	25	36	61*	38	45	83*
+	−	0	0	0	0	2	2*
−	−	24	36	60	5	32	37

The number of mice is shown for each histopathology group. Animals were killed at the age of 10 months. Pancreata and submandibular glands were removed and immersed in Bouin's fixative. Three serial sections (5- μ m thick) were stained with haematoxylin and eosin and scored for mononuclear cell infiltration. Examination of 22 C57BL/6J and 18 NZW 10-month-old parent mice showed no pancreatic inflammation and infrequent and minor salivary gland lesions (1.7 $\pm$ 3.1 foci in males and 1.9 $\pm$ 2.0 foci in females) when compared with NOD mice at the same age (5.3 $\pm$ 2.6 and 15.6 $\pm$ 5.4 foci in males and females, respectively). None of these F₂ animals became diabetic.

* $P < 10^{-9}$ for the paired comparisons test of mice with periinsulinitis only and those with sialitis only.

least a partly common molecular defect underlies both extrapancreatic and pancreatic inflammatory lesions in NOD mice.

To investigate this hypothesis in further detail, we studied the segregation of periinsulinitis and of submandibulitis taken as an example of extrapancreatic inflammation in two F₂ crosses between NOD and nondiabetic C57BL/6J and NZW mice. At 10 months of age, when all mice were killed, the two latter strains showed no pancreatic inflammation and negligible infiltrative lesions in their salivary glands, compared with those of NOD mice. The NZW genetic background was chosen because it contributes to the severe lupus-like syndrome of (NZB × NZW)F₁ hybrid mice, and particularly to their marked Sjögren's syndrome²⁰. Almost all F₂ animals with periinsulinitis were also scored positively for sialitis ($P < 10^{-9}$ for the paired comparisons test of groups of mice showing either trait; Table 1). This tight linkage was asymmetrical as a number of F₂ mice had sialitis without periinsulinitis. It strongly suggests overlapping genetic control for these two traits. Interestingly, few severe pancreatic lesions were seen in either cross (two and three mice with invasive insulinitis among (NOD × B6)F₂ and (NOD × NZW)F₂ animals, respectively). Thus, although the incidence of disease was greater in the (NOD × NZW)F₂ cross, there was

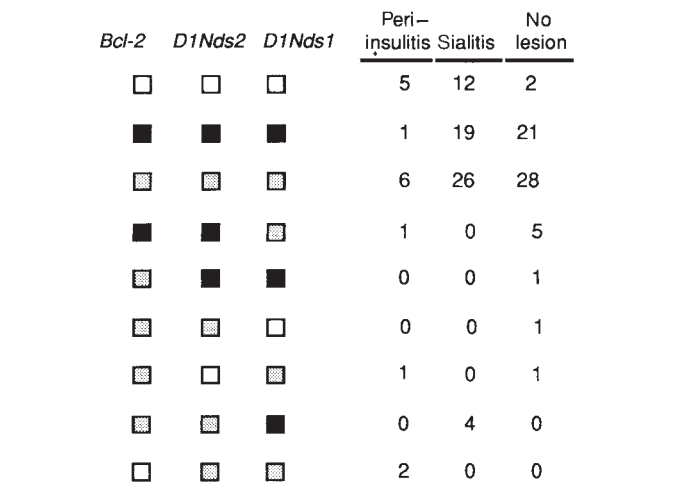


FIG. 1 Chromosome 1-linked genotypes of (NOD × B6)F₂ mice in the vicinity of the *Bcl-2* gene. On each row, the number of mice with the given haplotype is shown for each histopathology group. Symbols: open, homozygous *nod*; solid, homozygous *b6*; shaded, heterozygous.

TABLE 2 Association of histopathology with chromosome 1 in (NOD × B6)_F₂ mice

Chromosome	Locus (distance in cM)	Allele typing	Periinsulitis and sialitis	Histopathology		χ^2 value	P value
				Sialitis only	No lesion		
1	<i>D1Nds4</i> (3)	<i>nod/nod</i>	4	18	12	5.96	<0.00025
		<i>nod/b</i>	9	32	25		
		<i>b/b</i>	4	11	22		
1	<i>Bcl-2</i> (41)	<i>nod/nod</i>	8	12	2	21.24	
		<i>nod/b</i>	7	30	31		
		<i>b/b</i>	2	19	26		
1	<i>D1Nds2</i> (42)	<i>nod/nod</i>	7	12	3	16.44	
		<i>nod/b</i>	8	30	29		
		<i>b/b</i>	2	19	27		
1	<i>D1Nds1</i> (48)	<i>nod/nod</i>	5	12	3	14.30	
		<i>nod/b</i>	11	26	29		
		<i>b/b</i>	1	23	27		
1	<i>Crp</i> (72)	<i>nod/nod</i>	3	11	7	1.51	
		<i>nod/b</i>	10	31	35		
		<i>b/b</i>	4	19	17		
3	<i>Il-2</i> (32)	<i>nod/nod</i>	5	18	12	2.33	
		<i>nod/b</i>	7	30	29		
		<i>b/b</i>	5	13	18		
3	<i>D3Nds1</i> (53)	<i>nod/nod</i>	4	18	11	2.97	
		<i>nod/b</i>	9	32	31		
		<i>b/b</i>	4	11	17		
9	<i>Cypla2</i> (33)	<i>nod/nod</i>	7	13	11	5.63	
		<i>nod/b</i>	5	33	28		
		<i>b/b</i>	5	15	20		
11	<i>Csfgm</i> (30)	<i>nod/nod</i>	3	10	12	0.35	
		<i>nod/b</i>	9	34	31		
		<i>b/b</i>	5	17	16		
14	<i>Plau</i> (8)	<i>nod/nod</i>	4	21	14	1.99	
		<i>nod/b</i>	7	23	26		
		<i>b/b</i>	6	17	19		
15	<i>Myc</i> (18)	<i>nod/nod</i>	3	10	13	1.19	
		<i>nod/b</i>	11	35	32		
		<i>b/b</i>	3	16	14		
17	<i>H-2</i> (19)	<i>nod/nod</i>	9	9	8	19.23	<0.0005
		<i>nod/b</i>	5	45	36		
		<i>b/b</i>	3	7	15		

Haplotypes at the *H-2* complex were determined serologically with anti-*I-A* monoclonal antibodies MKD6 and 10.3.6. All other loci were typed by PCR-amplification of microsatellite variants as described²¹⁻²⁴. Genetic distances in centimorgans are from ref. 13. Independence of histopathology and of genotype at each locus was tested by calculating the value of the χ^2 distribution for each of the three possible genotypes in each histopathologic group (4 degrees of freedom).

no strong synergy between NOD and NZW genetic backgrounds to determine severe diabetes.

These segregation data, obtained by using two distinct crosses with two different strain combinations, confirm that sialitis is an important disease marker in the NOD mouse and suggest that periinsulinitis is part of a diffuse inflammatory process. It led to the delineation of three main histopathological groups comprising: (1) mice with both periinsulinitis and salivary gland inflammation; (2) mice with sialitis only; and (3) mice with neither of these two lesions. These three subgroups were considered separately to increase the resolution of genetic analysis.

To identify and to map susceptibility loci, we analysed the genomes of (NOD × B6)_F₂ animals by using polymerase chain reaction (PCR) amplification of polymorphic microsatellite repeats following the procedure pioneered by J. Todd in the mouse²¹⁻²⁴. Histopathology was most significantly associated with the *Bcl-2* locus in the middle region of chromosome 1 (Table 2; $P < 0.00025$). To confirm this finding, additional chromosome 1-linked loci were examined, which explains their overrepresentation in this panel. The most significant association remained with the *Bcl-2* marker. Genotype distribution at closely linked *D1Nds2* and *D1Nds1* loci was also markedly affected, whereas *D1Nds4* and *Crp* that lie further away from *Bcl-2* toward the centromere and the telomere, respectively, did not seem to be associated with disease.

A slight distortion of segregation with a decrease in homozygous *nod/nod* animals can be observed at all four loci map-

ping to the distal half of chromosome 1 ($0.01 < P < 0.05$ when compared with the theoretical 1:2:1 distribution at the *Bcl-2* locus) but not at the most proximal *D1Nds4* locus. There is no simple explanation for this finding. Mistyping is unlikely because the genotypes at these linked loci are concordant and in agreement with recombination distances calculated from other sources (Fig. 1; ref. 13, and our unpublished data on 450 [(NOD × B6) × NOD]BC1 mice). The interpretation of our results is nonetheless not affected as their significance is assessed by a χ^2 test for independence of genotype and of histopathology. Intriguingly, recent analysis of another NOD cross¹³ also reveals a segregation distortion at several loci, in particular at pathological *D3Nds1* and *D11Nds11* loci.

The incidence of periinsulinitis and of sialitis in each genotypic group at the *Bcl-2* locus indicates that these traits are both influenced by *nod* and *b* haplotypes on a codominant mode as the occurrence of lesions was increased in homozygous *Bcl-2^{nod/nod}* mice and decreased in *Bcl-2^{b/b}* mice (Table 3). High penetrance of sialitis in *Bcl-2^{nod/nod}* homozygous animals suggests that the *Bcl-2*-linked gene is a major determinant of this trait, at least in the homozygous state. By contrast, only 36% of *Bcl-2^{nod/nod}* homozygous mice had periinsulinitis, indicating that additional genes are involved in the occurrence of this lesion. Extensive analysis of the genomes of these mice is now in progress and should help to identify such genes.

Among other loci we tested, the *H-2* complex had much involvement. Sialitis, however, was less influenced by *H-2^{nod}*

TABLE 3 The contribution of *Bcl-2* linked *nod* haplotype to periinsulinitis and sialitis in (NOD × B6)F₂ mice

<i>Bcl-2</i> type	<i>nod/nod</i>	<i>nod/b</i>	<i>b/b</i>
Sialitis	90.9% (<i>P</i> < 0.0005)	54.4% NS	44.7% (<i>P</i> < 0.05)
Periinsulinitis	36.3% (<i>P</i> < 0.0005)	10.3% NS	4.2% (<i>P</i> < 0.05)

Differences in the frequencies of animals with the given *Bcl-2* genotype and those with other genotypes in each histopathology category are assessed by χ^2 analysis in a 2 × 2 contingency table, with Yates' correction when necessary. NS, not significant.

haplotype than periinsulinitis. Conversely, chromosomes 3 and 11 which were recently shown to be associated with invasive insulinitis and diabetes onset¹³ were not associated with the less severe histopathology seen in our cross. This indicates that distinct genes can be characterized at each of the successive steps in the progression of the disease as suggested¹².

The precise location and the identity of this new chromosome 1-linked susceptibility gene remain to be determined. Examination of recombinant chromosome 1 haplotypes suggests tentatively mapping it between *Bcl-2/D1Nds2* and *D1Nds1* (Fig. 1). It should be stressed that this locus was identified using NOD and C57BL/6J parent mice. These results need to be extended to other strain combinations. As for human type-1 diabetes, the existence of conserved syntenies between the

human and the murine genomes²⁵ points to a possible human homologue mapping to chromosomes 1, 2 or 18. □

Received 24 May; accepted 26 July 1991.

- Cahill, G. F. & McDevitt, H. O. *New Engl. J. Med.* **304**, 1454–1465 (1981).
- Gorsuch, A. N. *et al. Lancet* **ii**, 1363–1365 (1981).
- Gepts, W. & Lecompte, P. M. *Am. J. Med.* **70**, 105–115 (1981).
- Foulis, A. K. & Stewart, J. A. *Diabetologia* **26**, 456–461 (1984).
- Foulis, A. K., Liddle, C. N., Farquharson, M. A., Richmond, J. A. & Weir, R. S. *Diabetologia* **29**, 267–274 (1986).
- Signore, A., Pozzilli, P., Gale, E. A. M., Andreani, D. & Beverley, P. C. L. *Diabetologia* **32**, 282–289 (1989).
- Bedossa, P., Bendelac, A., Bach, J.-F. & Carnaud, C. *Eur. J. Immun.* **19**, 1947–1951 (1989).
- Sveigaard, A., Platz, P. & Ryder, L. P. *Immunol. Rev.* **70**, 193–218 (1983).
- Todd, J. A., Bell, J. I. & McDevitt, H. O. *Nature* **329**, 599–604 (1987).
- Hattori, M. *et al. Science* **231**, 733–735 (1986).
- Prochazka, M., Leiter, E. H., Serreze, D. V. & Coleman, D. L. *Science* **237**, 286–289 (1987).
- Wicker, L. S. *et al. J. exp. Med.* **165**, 1639–1654 (1987).
- Todd, J. A. *et al. Nature* **351**, 542–547 (1991).
- Asamoto, H. *et al. J. Jap. Diab. Soc.* **27**, 775–781 (1984).
- Miyagawa, J. *et al. Virchows Arch. Cell Pathol.* **51**, 215–225 (1986).
- Whaley, K., Williamson, J. & Chisholm, W. W. *Q. J. Med.* **42**, 279–304, 1973.
- Ludwig, H. *et al. Gut* **18**, 311–315 (1977).
- Okazaki, K., Tamura, S., Morita, M., Nishimori, I. & Yamamoto, Y. *Int. J. Pancreat.* **5**, 359–377 (1990).
- Goillot, E., Mutin, M. & Touraine, J.-L. *Clin. Immun. Immunopathol.* **59**, 462–473 (1991).
- Kessler, H. S. *Am. J. Path.* **52**, 671–685 (1968).
- Love, J. M., Knight, A. M., McAleer, M. A. & Todd, J. A. *Nucleic Acids Res.* **18**, 4123–4130 (1990).
- Hearne, C. M. *et al. Mammalian Genome* (in the press).
- Aitman, T. J., Hearne, C. M., McAleer, M. A. & Todd, J. A. *Mammalian Genome* (in the press).
- Cornall, R. J., Aitman, T. J., Hearne, C. M. & Todd, J. A. *Genomics* (in the press).
- Nadeau, J. H. & Reiner, A. H. in *Genetic Variants and Strains of the Laboratory Mouse* (eds Lyon, M. F. & Searle, A. G.) 506–536 (Oxford University Press, Oxford, 1990).

ACKNOWLEDGEMENTS. We thank J. Bacci for technical assistance, C. Carnaud for help in the initiation of this work and for discussion, M. Lévi-Strauss for reading the manuscript and J. Todd for PCR primers and for information about these primers before publication.

Type 1 diabetes in mice is linked to the interleukin-1 receptor and *Lsh/Ity/Bcg* genes on chromosome 1

Richard J. Cornall, Jan-Bas Prins, John A. Todd, Alison Pressey*, Nicole H. DeLarato*, Linda S. Wicker* & Laurence B. Peterson†

Nuffield Department of Surgery, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

* Autoimmune Diseases Research, † Department of Cellular and Molecular Pharmacology, Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065, USA

HUMAN type 1 (insulin-dependent) diabetes is a common autoimmune disease of the insulin-producing β cells of the pancreas which is caused by both genetic and environmental factors^{1,2}. Several features of the genetics and immunopathology of diabetes in nonobese diabetic (NOD) mice are shared with the human disease^{1,3,4}. Of the three diabetes-susceptibility genes, *Idd-1* (refs 5–7), *-3* and *-4* (ref. 4) that have been mapped in mice to date, only in the case of *Idd-1* is there any evidence for the identity of the gene product: allelic variation within the murine immune response *I-A β* gene and its human homologue *HLA-DQB1* correlates with susceptibility, implying that *I-A β* is a component of *Idd-1* (refs 5–11). We report here the mapping of *Idd-5* to the proximal region of mouse chromosome 1. This region contains at least two candidate susceptibility genes, the interleukin-1 receptor gene (*Il-1r1*; ref. 12) and *Lsh/Ity/Bcg* (refs 13–19), which encodes resistance to bacterial and parasitic infections and affects the function of macrophages.

Previously, we analysed the mouse genome for susceptibility genes unlinked to *Idd-1* in the major histocompatibility complex (MHC) on chromosome 17 and mapped *Idd-3* and *Idd-4* to chromosomes 3 and 11, respectively⁴. The chromosome 1 marker

D1Nds4 showed a positive association with diabetes that was not significant⁴. We have now extended our analysis using additional diabetic and nondiabetic mice and by typing four more chromosome 1 markers (Table 1).

We analysed the genotype frequencies of eight markers on chromosome 1 (Fig. 1 and Table 1) in 106 spontaneously diabetic and 190 nondiabetic female BC1 progeny (from the reciprocal backcrosses, (B10.*H-2^b* × NOD)F₁ × NOD and NOD × (B10.*H-2^b* × NOD)F₁; B10 is C57BL/10; ref. 4). Three markers were associated with overt disease, two randomly cloned microsatellites *D1Nds4* and *MIT-L2*, and a third, designated *Il-1r1**, which was generated using polymerase chain reaction (PCR) primers designed from the published *Il-1r1* sequence (Table 1). Additional data were obtained from BC1 progeny treated with cyclophosphamide.

Cyclophosphamide is an alkylating drug that kills dividing cells and increases the frequency of autoimmune diabetes in NOD mice²⁰. Its diabetogenic effects are observed only in NOD mice that have already developed insulinitis²⁰, and not in F₁ hybrids of NOD with diabetes-resistant strains²¹ or in the NOD *H-2^b* congenic strain^{22,23}. Cyclophosphamide-induced diabetes is not due to a direct toxic effect of the drug on β cells because spleen cells from cyclophosphamide-induced diabetics can transfer diabetes to nondiabetic recipients²¹. The immunopathology of cyclophosphamide-induced diabetes is indistinguishable from that of the spontaneous disease^{20,22,23}. Furthermore, the only associations of marker loci with the induced diabetes that we have found correspond to markers already associated with spontaneous type 1 diabetes (S. Ghosh, J.A.T., R.J.C., A.P., N.H.D., L.S.W., L.B.P., manuscript in preparation). On the basis of these data, it is assumed that cyclophosphamide increases the frequency of diabetes by increasing the penetrance of certain NOD susceptibility alleles. Of 373 BC1 progeny treated with cyclophosphamide at ages ranging from 165 to 300 days, 81/293 (28%) female and 22/80 (28%) male BC1 progeny became diabetic. DNA from these diabetics was analysed and genotype frequencies of the markers compared with 98 of the cyclophosphamide-treated, nondiabetics. *D1Nds4*, *Il-1r1**, *MIT-L2*, *MIT-L31*, *Bcl-2* and *D1Nds1* were

TABLE 1 The association of chromosome 1 with spontaneous and cyclophosphamide-induced type 1 diabetes

Chromosome location (cM)	Locus	Genotypes	Untreated mice			Cyclophosphamide-treated mice			Combined		
			Diabetics	Non-diabetics	$\chi^2 > 4$	Diabetics	Non-diabetics	$\chi^2 > 4$	Diabetics	Non-diabetics	$\chi^2 > 4$
1(7)	<i>D1Nds4</i>	nb	42	108	8.1	35	52	7.4	77	160	17.0
		nn	64	82		68	46		132	128	
1(8)	<i>Il-1r1*</i>	nb	42	107	7.6	35	51	6.7	77	158	15.8
		nn	64	83		68	47		132	130	
1(24)	<i>MIT-L2</i>	nb	42	108	8.1	37	56	9.1	79	164	17.8
		nn	64	82		66	42		130	124	
1(36)	<i>MIT-L31</i>	nb	44	98		40	57	7.5	84	155	9.0
		nn	62	92		63	41		125	133	
1(41)	<i>Bcl-2</i>	nb	47	95		41	53	4.1	88	148	4.2
		nn	59	95		62	45		121	140	
1(46)	<i>D1Nds1</i>	nb	53	94		43	55	4.2	96	149	
		nn	53	96		60	43		113	139	
1(66)	<i>Fcgr2</i>	nb	59	86		53	46		112	132	
		nn	47	104		50	52		97	156	
1(67)	<i>Crp</i>	nb	59	88		54	45		113	133	
		nn	47	102		49	53		96	155	

Markers were ordered by minimizing double recombinants in the typing data from 497 BC1 progeny. These included the 194 progeny analysed previously⁴ with the following additional mice: 9 spontaneous diabetics (including 6 males), 93 non-cyclophosphamide-treated nondiabetics (female), 103 cyclophosphamide-treated overt diabetics (including 22 males) and 98 cyclophosphamide-treated nondiabetics (female). Map distances were calculated from recombination frequencies using Kosambi's formula and were consistent with published distances^{12,16-18}. Microsatellite markers *D1Nds4* (ref. 4), *Bcl-2* (ref. 27), *D1Nds1* (ref. 28) and *Crp* (ref. 29) have been described. *MIT-L2* and *MIT-L31* are randomly cloned microsatellites from W. Dietrich and E. Lander (unpublished data). *Il-1r1** primers were, 5'-TGAGCTGTCTGTCATTCTTG-3' and 5'-CCGTGCATTTTATTGGAGTAAG-3' at positions -264 and 107 of the published *Il-1r1* complementary DNA sequence²⁶, respectively, giving an 820-bp product with DNA from B10, C57BL/6J, CBA/J, C3H/HeJ, A/J, C57L/J, C57BR/cdJ, BALB/cByJ, SJL/J and *Mus spretus* (Pasteur Institute), but not with DNA from NOD, NON, SWR/J, DBA/2J and PL/J (PCR conditions: 2 mM Mg²⁺, 55 °C annealing temperature, 3 min extension; data not shown). Primers 5'-TTAGGTGATGATAATGATGAAT-3' and 5'-CGTAAGTCCACATTACCTT-3' were designed from the *Fcgr2* sequence³⁰, and used in PCR at 55 °C annealing temperature and 3 mM Mg²⁺ to give a product size of ~90 bp. Genotype frequency differences were tested by χ^2 in 2 × 2 contingency tables. Animals were monitored for elevations in urinary glucose using Tes-Tape (Eli Lilly) and were classified as diabetic after producing consistent Tes-tape values >2⁺. Diagnosis of type-1 diabetes was confirmed by histology of the pancreas^{4,22,23}. Pancreata were fixed in neutral-buffered formalin and processed for paraffin embedding. Tissue sections (4 µm) were stained with haematoxylin and eosin. Histology of mice classified as diabetic with high levels of glucose in the urine showed massive destruction of islets. Nondiabetic mice of >165 days old received 200 mg kg⁻¹ of cyclophosphamide intraperitoneally on days 0 and 14 and were monitored for diabetes until day 28. nn, NOD homozygous; nb, heterozygous.

associated with induced diabetes (Table 1). There were no significant differences in the genotype frequencies between the non-treated and the treated groups, so the data were combined, giving highly significant associations for *D1Nds4*, *Il-1r1** and *MIT-L2* with overt disease ($\chi^2 > 13.8$; $P < 0.001$; one degree of freedom; Table 1). A χ^2 of 13.8 is considered significant evidence

for linkage⁴. The gene (or genes) responsible for this diabetogenic effect is designated *Idd-5*. The effect of the NOD allele of *Idd-5*, like *Idd-1*, -3 and -4 (ref. 4), is not recessive because there are 27/106 (25%) spontaneous diabetics and 23/103 (22%) cyclophosphamide-induced diabetics that are heterozygous at *MIT-L31*, *MIT-L2*, *Il-1r1** and *D1Nds4*.

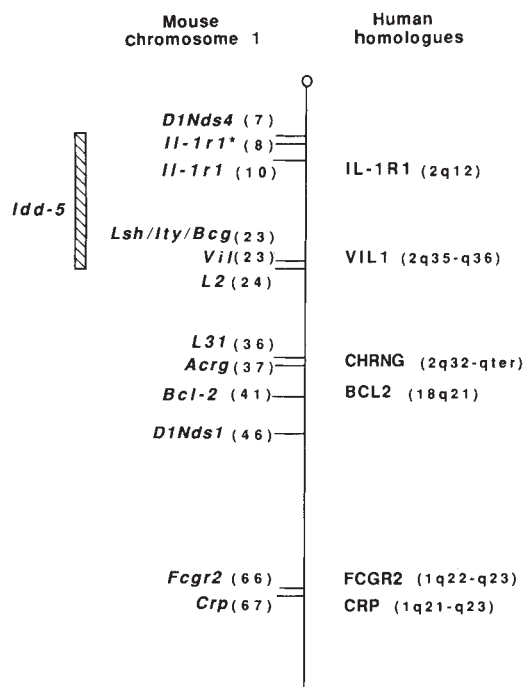


FIG. 1 Composite genetic map of mouse chromosome 1 with the human homologues and their approximate chromosomal locations using data from Table 1 and the literature^{12,16-18}. The markers used and mapped in this study are detailed in Table 1. *Bcl-2* was used as the anchor locus for the map. The location of *Idd-5* (stippled bar) is shown over the minimum region of chromosome 1, between *D1Nds4* and *MIT-L2*, in which there is at least 95% probability of linkage to overt diabetes.

TABLE 2 The association of chromosome 1 with insulinitis in nondiabetic mice not treated with cyclophosphamide

Chromosome location (cM)	Locus	Genotype	Genotype frequencies insulinitis grade			$\chi^2 > 4$	P value
			0, 1, 2, 2b	3	3b, 3b*, 4*		
1(7)	<i>D1Nds4</i>	nb	69	9	30	14.4	<0.001
		nn	32	5	45		
1(8)	<i>Il-1r1*</i>	nb	69	9	29	15.8	<0.001
		nn	32	5	46		
1(24)	<i>MIT-L2</i>	nb	68	10	30	14.4	<0.001
		nn	33	4	45		
1(36)	<i>MIT-L31</i>	nb	63	10	25	16.9	<0.001
		nn	38	4	50		
1(41)	<i>Bcl-2</i>	nb	59	10	26	12.5	<0.01
		nn	42	4	49		
1(46)	<i>D1Nds1</i>	nb	54	10	30	6.0	<0.05
		nn	47	4	45		
1(66)	<i>Fcgr2</i>	nb	41	8	37		
		nn	60	6	38		
1(67)	<i>Crp</i>	nb	41	8	39		
		nn	60	6	36		

Histological analysis was performed on pancreata from 210–330-day-old nondiabetic BC1 mice, not treated with cyclophosphamide. Histology was scored in the following categories: 0, no abnormal lymphocyte accumulations; 1, periductal/perivascular lymphocytic infiltration; 2 and 2b, periislet infiltration, no insulinitis; 3, very mild insulinitis in some islets with no reduction in islet mass; 3b, extensive insulinitis with significant islet mass remaining; 3b*, extensive insulinitis with significant reduction in islet mass; 4*, as 3b*, but only residual islets remaining. Genotype frequency differences were tested by χ^2 in 2×3 contingency tables (2 degrees of freedom). The total numbers of homozygous and heterozygous mice were not equal because the spontaneous diabetics were excluded from this analysis. nn, NOD homozygous; nb, heterozygous.

Idd-5 also influences the development of insulinitis in the 190 nondiabetic, non-treated BC1 mice (Table 2). NOD homozygosity at *D1Nds4*, *Il-1r1**, *MIT-L2* and *MIT-L31* was associated with an increase in the frequency of insulinitis ($P < 0.001$; two degrees of freedom). Three genes have now been associated with insulinitis, *Idd-1* on chromosome 17 (refs 5–7), *Idd-3* on chromosome 3 (ref. 4), and *Idd-5* on chromosome 1.

Our aim is to identify the functions of the non-MHC genes in type 1 diabetes. To this end, we have shown that *Idd-5* is linked to allelic variation in an *Il-1r1*-related sequence, and is also in the same region of chromosome 1 as the villin gene (*Vil*; Fig. 1), which in other strain combinations is within 1 centimorgan (cM) of *Lsh/Ity/Bcg* (refs 15–18). *Il-1r1* is a candidate susceptibility gene for type-1 diabetes because interleukin-1 is an important cofactor for T-cell activation and interleukin-1 administration *in vivo* can prevent NOD diabetes²⁴. The effects of interleukin-1 on type 1 diabetes are complex because this cytokine is toxic to β cells after culture *in vitro*²⁵. PCR analysis of genomic DNA using primers to the 5' region of the complementary DNA sequence of *Il-1r1* (ref. 26) gives a product of 820 base pairs in B10 but not in NOD (Table 1; and data not shown). In our backcrosses this sequence variation mapped to within 2 cM of the previously determined position of *Il-1r1* (ref. 12; Table 1). The precise nature of this variant and its relationship to type-1 diabetes are unknown.

Lsh/Ity/Bcg, which is believed to be a single gene, determines susceptibility to *Leishmania donovani* (*Lsh*), *Salmonella typhimurium* (*Ity*) and *Mycobacterium bovis* Bacille Calmette Guérin (*Bcg*) and has also been mapped to proximal mouse chromosome 1 (Fig. 1)^{13–18}. The *Lsh/Ity/Bcg* gene, which has yet to be cloned, can also be considered as a candidate for *Idd-5* because it influences macrophage activation^{13–15}. It is known that macrophages are required for the development of diabetes in NOD mice¹⁹. Furthermore, the NOD strain is resistant to infection with *L. donovani* (P. Kaye, A. Cooke, A. M. Wattie, T. Lund and J. M. Blackwell, manuscript in preparation), whereas B10 is susceptible¹⁵.

In the diabetics, the segregation of the chromosome 1 marker, *MIT-L2* with the most closely linked markers to disease on chromosomes 3 (*D3Nds1*) and 11 (*D11Nds1*) does not differ significantly from random expectations (data not shown). The

data indicate that the NOD alleles of *Idd-3*, *-4* and *-5* might act in a predominantly additive way with an epistatic component (S. Ghosh *et al.*, manuscript in preparation).

The homologue of *Lsh/Ity/Bcg* may map to human chromosome 2q (Fig. 1) because preliminary data suggest that susceptibility to tuberculosis and leprosy could be linked to human chromosome 2q (ref. 14). The *Il-1r1* homologue maps to human 2q (ref. 12), and is part of a 30-cM region of homology that extends from *Il-1r1* to *Acrg* on mouse chromosome 1 (refs 12, 18; Fig. 1). The human homologues of the 13 genes that have so far been mapped to this interval on mouse chromosome 1 have all been located on human chromosome 2q (ref. 18), suggesting that this region of chromosome 2 may contain a gene that influences human type 1 diabetes. □

Received 13 June; accepted 6 August 1991.

- Castano, L. & Eisenbarth, G. S. *A. Rev. Immun.* **8**, 647–679 (1990).
- Risch, N. *Am. J. hum. Genet.* **40**, 1–14 (1987).
- Makino, S. *et al. Expl. Anim.* **29**, 1–13 (1980).
- Todd, J. A. *et al. Nature* **351**, 542–547 (1991).
- Hattori, M. *et al. Science* **231**, 733–735 (1986).
- Prochazka, M., Leiter, E. H., Serreze, D. V. & Coleman, D. L. *Science* **237**, 286–289 (1987).
- Wicker, L. S. *et al. J. exp. Med.* **165**, 1639–1654 (1987).
- Todd, J. A., Bell, J. I. & McDewitt, H. O. *Nature* **329**, 599–604 (1987).
- Miyazaki, T. *et al. Nature* **345**, 722–724 (1990).
- Slattery, R. M. *et al. Nature* **345**, 724–726 (1990).
- Lund, T. *et al. Nature* **345**, 727–729 (1990).
- Copeland, N. G. *et al. Genomics* **9**, 44–50 (1991).
- Blackwell, J. M. *et al. Immun. Lett.* (in the press).
- Schurr, E. *et al. Immun. Today* **12**, 42–45 (1991).
- Skamene, E. *et al. Nature* **297**, 506–509 (1982).
- Malo, D. *et al. Genomics* **10**, 356–364 (1991).
- Mock, B. *et al. Genomics* **7**, 57–64 (1990).
- Hillyard, A. L., Doolittle, D. P., Davisson, M. T. & Roderick, T. H. *Locus Map of the Mouse with Comparative Points of Human on Mouse* (Jackson Laboratory, Bar Harbor, Maine, 1991).
- Hutchings, P. *et al. Nature* **348**, 639–642 (1990).
- Harada, M. & Makino, S. *Diabetologia* **27**, 604–606 (1984).
- Yasunami, R. & Bach, J.-F. *Eur. J. Immun.* **18**, 481–484 (1988).
- Wicker, L. S., DeLarato, N. H., Pressey, A. & Peterson, L. B. in *Proc. P & S Biomed. Sci. Symp.: Molecular Mechanisms of Immunological Self-recognition* (eds Vogel, H. J. & Alt, S. W.) (Academic, New York, in the press).
- Wicker, L. S., DeLarato, N. H., Pressey, A. & Peterson, L. B. in *Frontiers in Diabetes Research II: Lessons from Animal Diabetes III* (ed. Shafritz, E.) 67–72 (Smith-Gordon, London, 1991).
- Jacob, C. O., Aiso, S., Michie, S. & McDewitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **87**, 968–972 (1990).
- Mandrup-Poulsen, T. *et al. Curr. Topics Microbial. Immun.* **164**, 169–193 (1990).
- Sims, J. E. *et al. Science* **241**, 585–589 (1988).
- Love, J. M., Knight, A. M., McAleer, M. A. & Todd, J. A. *Nucleic Acids Res.* **18**, 4123–4130 (1990).
- Cornall, R. J., Aitman, T. J., Hearne, C. M. & Todd, J. A. *Genomics* **10**, 874–881 (1991).

29. Hearne, C. M. *et al.* *Mamm. Genome* **1**, 273–282 (1991).
 30. Hogarth, P. M. *et al.* *J. Immun.* **146**, 369–377 (1991).

ACKNOWLEDGEMENTS. We thank colleagues H.-J. Garchon, A. Cooke, W. Dietrich and E. Lander for sharing data before publication, W. Dietrich and E. Lander for microsatellites *MIT-L2* and *MIT-L31*, and J. Love, N. Rodrigues, S. Ghosh, C. Hearne, M. McAleer, S. Bain, P. Neumann and T. Aitman for help. This work was supported by the Wellcome Trust, the Medical Research Council (as part of the UK Human Genome Mapping Project), the Juvenile Diabetes Foundation, Lilly Industries Limited (UK) and the British Diabetes Association. R.J.C. is an MRC Training Fellow.

Inhibition of early atherogenesis in transgenic mice by human apolipoprotein AI

Edward M. Rubin^{*†}, Ronald M. Krauss^{*},
 Elizabeth A. Spangler^{*}, Judy G. Verstuyft^{*}
 & Shirley M. Clift^{*}

^{*} Division of Research Medicine and Radiation Biophysics and

[†] Division of Cell and Molecular Biology, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, USA

EPIDEMIOLOGICAL surveys have identified a strong inverse relationship between the amount in the plasma of high density lipoproteins (HDL), apolipoprotein AI (ApoA-I), the major protein component of HDL, and the risk for atherosclerosis in humans^{1,2}. It is not known if this relationship arises from a direct antiatherogenic effect of these plasma components or if it is the result of other factors also associated with increases in ApoA-I and HDL levels. Because some strains of mice are susceptible to diet-induced formation of preatherosclerotic fatty streak lesions, and because of available techniques for the genetic manipulation of this organism, the murine system offers a unique setting in which to investigate the process of early atherogenesis. To test the hypothesis that induction of a high plasma concentration of ApoA-I and HDL would inhibit this process, we studied the effects of atherogenic diets on transgenic mice expressing high amounts of human ApoA-I. We report that transgenic mice with high plasma ApoA-I and HDL levels were significantly protected from the development of fatty streak lesions.

Atherosclerosis susceptibility in mice is a characteristic of particular inbred strains, and results from the interaction of several genes^{3–5}, none of which map to the murine ApoA-I gene. The degree of susceptibility differs from strain to strain, as assessed by the area of lipid staining fatty streak lesions formed in a defined segment of the aorta after exposure to an atherogenic diet^{3,6,7}. The gradation of susceptibility among different inbred strains of mice is illustrated by a comparison of C57BL/6, BALB/c and C3H mice. Animals of strain C57BL/6 consistently develop lesions after a limited (14–18 weeks) exposure to an atherogenic diet, whereas BALB/c and C3H mice remain free of lesions. A year-long exposure of these strains to the diet results in the appearance of complex atheromatous plaques in C57BL/6, occasional fatty streaks in BALB/c and no lesions in C3H mice⁷.

Strain C57BL/6 differs from strains more resistant to atherosclerosis by a reduction in plasma HDL when fed an atherogenic diet^{3,6,8}. Because of the observed correlation between plasma HDL levels and atherosclerosis in humans, this reduction in HDL in C57BL/6 mice is postulated to be responsible for the extreme susceptibility of this strain to the development of diet-induced fatty streak lesions. To determine if high amounts of ApoA-I and HDL could alter the atherogenic susceptibility of C57BL/6 mice, we measured the extent of diet-induced fatty streak lesion formation in transgenic C57BL/6 mice expressing human ApoA-I (ref. 9). The effects of two high-fat atherogenic diets were studied in human ApoA-I transgenic mice and in age- and sex-matched C57BL/6 control animals. Transgenic and control animals were housed in the

same cages and fed either of the two atherogenic diets (Table 1). After exposure to the diets, the animals were killed, blood was collected for apolipoprotein and lipid measurements, and each aorta examined for lipid staining fatty streak lesions.

The non-HDL cholesterol fraction was higher in all animals fed either of the high-fat diets, but did not differ significantly between transgenic and control mice when fed the same diet. The plasma HDL-cholesterol concentration of the transgenic mice was about twice that of non-transgenic controls on low-fat and on both high-fat diets (Table 1). The total plasma ApoA-I concentration of the transgenic mice was also more than twofold higher than that of control mice on all three diets. The increase in total ApoA-I in the transgenic mice reflects a very high plasma concentration of human ApoA-I, and occurs in the context of a reduction in the amount of murine ApoA-I in the plasma compared with that of non-transgenic control mice.

The size distribution of very low density and low density lipoprotein particles did not differ significantly between any of the groups of mice, whereas large differences were observed in their HDL particle size profiles (Fig. 1). There is a single major population of HDL particles in plasma from control mice (Fig. 1a, b), whereas in transgenic mouse plasma there are three major size subclasses of HDL, one similar in size to the single HDL size subclass of the non-transgenic mice, and a larger and a smaller HDL size subclass (Fig. 1c, d). There were small increases in the particle size of the HDL subclasses on high-fat diets in both groups, most notably in the largest HDL species of the transgenic animals. The characteristic HDL profile of the female ApoA-I transgenic mice used in this study differs from that previously reported for male ApoA-I transgenic mice⁹. As in the males, however, human ApoA-I is the major apolipoprotein associated with each of the HDL-size subclasses in the female transgenics (data not shown).

Aortic sections for the analysis of fatty streak lesions (Fig. 2) were prepared using the method described by Paigen *et al.*¹⁰. The quantitative aspects of the analysis were altered in the following ways to provide a more accurate statistical representation of the data. In their studies, Paigen *et al.*¹⁰ assume that

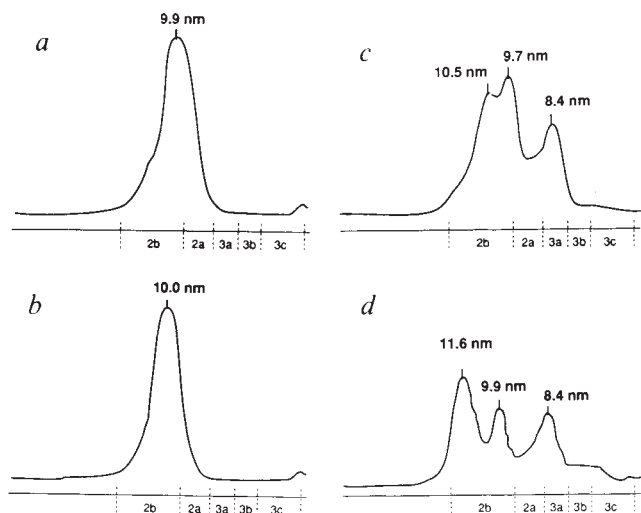


FIG. 1 Distribution by size of HDL particles. Plasma lipoproteins (density $<1.210 \text{ g ml}^{-1}$) were isolated from a 4-month-old female. **a**, C57BL/6 mouse fed a low-fat chow diet; **b**, C57BL/6 mouse fed high-fat diet A; **c**, ApoA-I transgenic C57BL/6 mouse fed a low-fat chow diet; and **d**, ApoA-I transgenic C57BL/6 mouse fed a high-fat Diet A. The lipoprotein fractions were subjected to nondenaturing electrophoresis in 4–30% polyacrylamide gradient gels¹⁷. Computer assisted scanning densitometry using coelectrophoresed molecular size calibrators was used to facilitate determination of lipoprotein particle sizes according to Nichols *et al.*²⁰.

TABLE 1 Amount of cholesterol and ApoA-I in plasma in transgenic and control mice fed low- and high-fat diets

	Mouse chow			Diet A			Diet B		
	Cholesterol (mg dl ⁻¹)			Cholesterol (mg dl ⁻¹)			Cholesterol (mg dl ⁻¹)		
	HDL	Non-HDL	Total	HDL	Non-HDL	Total	HDL	Non-HDL	Total
Controls	45 ± 2	9 ± 2	54 ± 3	36 ± 2	89 ± 7	125 ± 7	35 ± 4	121 ± 12	156 ± 13
Transgenics	82 ± 9*	18 ± 4*	100 ± 9*	85 ± 6*	70 ± 12	155 ± 17	88 ± 6*	169 ± 29	257 ± 31*
	ApoA-I† (mg dl ⁻¹)			ApoA-I† (mg dl ⁻¹)			ApoA-I† (mg dl ⁻¹)		
	Total	Human	Mouse	Total	Human	Mouse	Total	Humans	Mouse
Controls	75 ± 4	0	75 ± 4	90 ± 6	0	90 ± 6	75 ± 6	0	75 ± 6
Transgenics	212 ± 11*	201 ± 10*	11 ± 6*	222 ± 12*	145 ± 6*	35 ± 12*	178 ± 7*	147 ± 8*	13 ± 2*

Each diet study was done on groups of transgenic and control mice housed in the same cages and administered the diets at the same time. Control animals were sex- and age-matched C57BL/6 mice derived from the same vendor used in creating the ApoA-I transgenic founder animal. The low-fat diet was Purina laboratory mouse chow (5001) which contains 4.5% (w/w) animal fat <0.03% (w/w) cholesterol, no sodium cholate and no casein. High-fat diet A (ref. 21) contains 15% (w/w) fat (primary fat source is dairy butter), 1% cholesterol (w/w), 0.5% sodium cholate (w/w) and 20% casein (w/w). High-fat diet B (ref. 21) contains 15% (w/w) fat (primary fat source is cocoa butter), 1.25% (w/w) cholesterol, 0.5% (w/w) sodium cholate and 7.5% (w/w) casein. The low-fat diet measurements were done on 10 transgenic and 10 control mice at 10 weeks of age, the diet A measurements on 5 transgenic and 15 control mice at 28 weeks of age after 18 weeks on this diet, and the diet B measurements on 7 transgenic and 12 control mice at 24 weeks of age after 14 weeks on this diet. Total cholesterol concentration was determined by the procedure of Rudel and Morris²², whereas HDL-cholesterol was determined after the removal of non-HDL lipoproteins by selective precipitation with polyethylene glycol²³. Plasma ApoA-I was quantified by radial immunodiffusion assays done with anti-human and anti-mouse Apo-I antibodies as previously described⁹. (The standard error of the mean is given for each value.)

† Each value is the mean of the sum of murine and human ApoA-I.

* Significant differences between transgenic and control mice at $P < 0.05$.

lesions in different aortic sections are independent events, and assess the mean lesion area per section for a group of animals. To reflect the likelihood that lesions in different aortic sections in a given animal are connected, we have assessed the mean lesion area per section on a per animal basis for each group of animals. In addition, because the lesion area in the most proximal aortic section defined by Paigen *et al.*¹⁰ was much more variable than that of the remaining sections, we excluded the proximal section from our analysis. We find that the transgenic mice have a much smaller fatty streak lesion area per section on both atherogenic diets tested (Fig. 3). Expression of the human ApoA-I transgene resulted in complete protection from the development of fatty streak lesions in the transgenic animals on high-fat diet A. The more atherogenic diet, diet B, resulted in almost a sevenfold reduction in lesion area in transgenic mice compared with control animals on the same diet (551 μm^2 versus 3,890 μm^2).

The increase in total cholesterol after consumption of either high-fat diet was primarily the result of an increase in the non-HDL cholesterol fraction (Table 1). Despite the lack of a significant difference in the amount of non-HDL cholesterol between transgenic and control mice, the transgenic had a smaller fatty streak lesion area on both atherogenic diets. The transgenic mice also differed from the controls in the composition and size distribution of their HDL. In humans particular size subpopulations of HDL have been associated with a reduced risk for atherosclerosis^{11,12}. It is thus possible that alterations in the physical and chemical properties of HDL, as well as increased HDL plasma concentration, may contribute to the resistance of the transgenic mice to the development of diet-induced fatty streak lesions.

The reduction in the plasma concentration of murine ApoA-I in the transgenic mice is puzzling. We have previously demonstrated that this reduction is not associated with a decrease in

the amount of murine ApoA-I messenger RNA (ref. 9) whereas Walsh *et al.*¹³ have shown that human and murine ApoA-I are catabolized at similar rates after infusion into human ApoA-I transgenic mice suggesting that lipoproteins containing either murine or human ApoA-I are cleared at similar rates. Further studies will be necessary to elucidate the mechanism responsible for the decrease in murine ApoA-I plasma levels in the transgenic mice, and to determine if this process contributes to the resistance of these animals to diet-induced fatty streak lesion formation.

The plasma triglyceride amounts of human ApoA-I transgenic mice previously described^{9,13} and from this study (data not

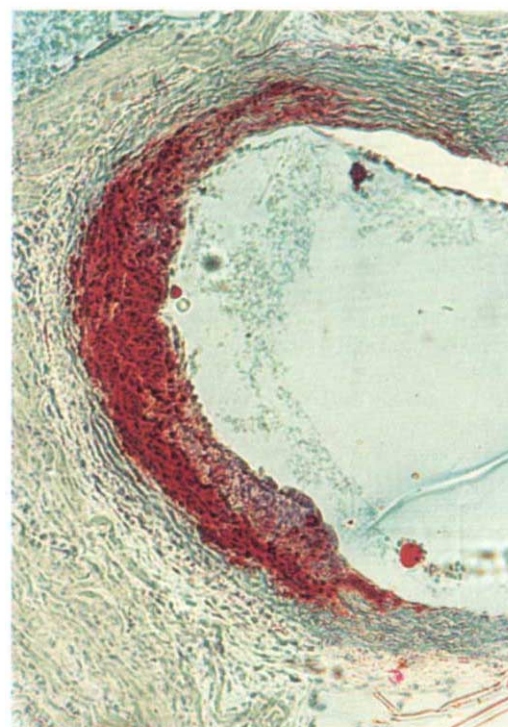


FIG. 2 Proximal aortic section from a C57BL/6 mouse fed the dairy butter diet, stained with Oil Red O and haematoxylin, and counterstained with light green. A large, red-stained preatherosclerotic fatty streak has invaded the media of this vessel. The heart and attached aorta were fixed and the sections prepared as previously described¹⁰. Magnification is $\times 200$.

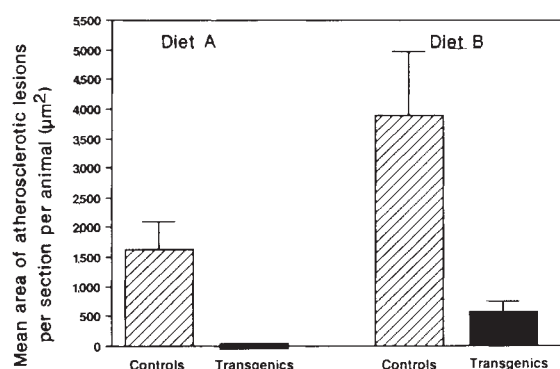


FIG. 3 Mean area of atherosclerotic lesions per section per animal for transgenic and control mice on high-fat diets. We quantitated four 10 µm aortic sections each separated by 80 µm. The first and most proximal section to the heart was taken 80 µm distal to the point where the aorta first becomes rounded. The area of Oil Red O staining fatty streak lesions in each section was determined using a calibrated eyepiece, and the mean lesion area per section per animal was calculated for each group of animals. Female mice were placed on the two high-fat diets at 10 weeks of age. After 18 weeks on diet A, 5 transgenic mice and 15 control mice were assayed for lesions. After 14 weeks on diet B, 7 transgenic mice and 12 control mice were assayed. The coded slides were examined blind in two separate analyses by the same examiner and gave consistent results ($R > 0.92$). The error bars depict the s.e.m. for each study. Significant difference between transgenic and control mice at $P < 0.05$.

shown) do not differ significantly from control animals. In humans, there is generally an inverse correlation between plasma HDL amounts and triglyceride amounts. The lack of a significant change in triglycerides in ApoA-I transgenic mice with high ApoA-I and HDL levels is consistent with evidence suggesting that this relationship is due to primary alterations in triglyceride and not HDL metabolism^{14,15}.

The results of our study are consistent with human^{1,2,16} and primate¹⁷ studies showing a strong inverse association between ApoA-I and HDL levels and the development of atherosclerotic lesions. The relationship between the fatty streak lesions measured in this study and mature atherosclerotic lesions which can form in mice after a longer dietary exposure^{6,7} has not been examined. In humans a precursor-product relationship seems to exist between these two types of lesions in the coronary arteries whereas a less defined relationship has been noted for other locations, including the aorta^{18,19}. The results of our study suggest that a high plasma concentration of human ApoA-I in transgenic mice has a direct inhibitory effect on the early stages of atherogenesis. Long-term studies will be necessary to assess the impact of ApoA-I on the formation of mature atherosclerotic plaques. □

Received 19 April; accepted 6 August 1991.

- Gordon, D. & Rifkin, B. *New Engl. J. Med.* **321**, 1311–1316 (1989).
- Gordon, D. et al. *Circulation* **79**, 8–15 (1989).
- Paigen, B., Mitchell, D., Reue, K., Morrow, A. & LeBoeuf, R. *Proc. natn. Acad. Sci., U.S.A.* **84**, 3763–3767 (1987).
- Paigen, B., Nesbitt, M. N., Mitchell, D., Albee, D. & LeBoeuf, R. *Genetics* **122**, 163–168 (1989).
- Stewart-Phillip, J., Lough, J. & Skomene, E. *Clin. Invest. Med.* **12**, 121–126 (1989).
- Ishida, B. & Paigen, B. in *Genetic Factors and Atherosclerosis: Approaches and Model Systems* (eds Lusis, A. J. & Sparkes, R. S.) 198–222 (Karger, Basel, 1989).
- Paigen, B., Ishida, B., Verstuyft, J., Winters, R. & Albee, D. *Arteriosclerosis* **10**, 316–323 (1990).
- LeBoeuf, R. C. et al. *J. Lipid Res.* **31**, 9–101 (1990).
- Rubin, E. M., Ishida, B. Y., Clift, S. M. & Krauss, R. M. *Proc. natn. Acad. Sci., U.S.A.* **88**, 434–438 (1991).
- Paigen, B., Morrow, A., Holmes, P. A., Mitchell, D. & Williams, R. A. *Atherosclerosis* **68**, 231–240 (1987).
- Miller, N. E. *Am. Heart J.* **113**, 589–597 (1987).
- Cheung, M., Brown, B., Wolf, J. A. & Albers, J. J. *Lipid Res.* **32**, 383–394 (1991).
- Walsh, A., Ito, Y. & Breslow, J. L. *J. Biol. Chem.* **264**, 6488–6494 (1989).
- Patsch, J., Karlin, L., Scott, L., Smith, L. C. & Gotto, A. *Proc. natn. Acad. Sci., U.S.A.* **80**, 1435–1449 (1983).
- Brinton, E., Eisenberg, S. & Breslow, J. L. *J. Clin. Invest.* **87**, 536–544 (1991).
- Miller, N. E. *Lipids* **13**, 914–919 (1978).
- Sorci-Thomas, M. et al. *J. Biol. Chem.* **263**, 5183–5189 (1988).
- Ross, R. *New Engl. J. Med.* **314**, 488–491 (1986).

- McGill, H. C. *Arteriosclerosis* **4**, 443–451 (1984).
- Nichols, A. V., Krauss, R. M. & Musliner, T. A. *Meth. Enzym.* **128**, 417–431 (1986).
- Nishina, P. M., Verstuyft, J. & Paigen, B. *J. Lipid Res.* **31**, 859–869 (1990).
- Rudel, L. L. & Morris, M. D. *J. Lipid Res.* **14**, 364–371 (1973).
- Izzo, C., Grillo, F. & Murador, E. *Clin. Chem.* **27**, 371–378 (1981).

ACKNOWLEDGEMENTS. We thank P. Nishina and B. Paigen for sharing information about murine diets, J. Bristow and P. Williams for discussion, and P. Blanche and B. Ishida for assistance with the lipoprotein and immunoassays. This research was funded by the National Dairy Promotion Research Board (E.M.R. and R.M.K.). E.A.S. was supported by an NRSA Training Grant. This work was conducted at the Lawrence Berkeley Laboratory.

Expression pattern of zebrafish *pax* genes suggests a role in early brain regionalization

Stefan Krauss*, Terje Johansen†, Vladimir Korzh* & Anders Fjose*‡

* Molecular Genetics Group and † Biotechnology Group, Department of Microbiology, Institute of Medical Biology, University of Tromsø, N-9000 Tromsø, Norway

IN vertebrates the developing hindbrain is organized in segmental units¹. These units provide the primary grid for differentiation and axonal outgrowth^{1,2}. In the more anterior regions of the brain, however, the subdivisions remain more controversial. Cellular and molecular studies of the embryonic brain in lower vertebrates such as the zebrafish, *Brachydanio rerio*, may reveal remnants of such subdivisions. We have isolated complementary DNA clones for two zebrafish *pax* genes related to *Drosophila* and mouse paired-box-containing segmentation genes^{3–10}. The expression of these two genes is confined to specific regions in the embryonic forebrain and midbrain. Strikingly, the borders of expression of the two *pax* genes coincide with morphological landmarks corresponding to the primary axon tracts that are generated in the embryonic brain a few hours after the initiation of expression of these genes^{11,12}.

We have started a search for zebrafish developmental genes that are expressed in the anterior brain. Complementary DNA clones for two such genes, *pax[zf-a]* and *pax[zf-b]*, were isolated (Fig. 1) and their expression patterns in the embryonic brain were investigated by *in situ* hybridization. The derived amino-acid sequences of both cDNAs (Fig. 1b) include the paired-box^{3,4}, a conserved domain capable of sequence-specific DNA-binding¹³. The *pax[zf-a]* (Fig. 1a) gene contains in addition a paired-homeobox³. Among the murine *pax* genes reported^{5–10}, only *Pax3* and *Pax7* contain both a paired-box and a homeobox. From Fig. 1b it can be seen that the C-terminal 89-amino-acid portion of the paired-box found in *pax[zf-a]* is only 66.7% and 62.2% identical to *Pax7* and *Pax3*, respectively. Moreover, the homeobox of *pax[zf-a]* is rather divergent compared with the five previously reported paired-type homeoboxes (65.6% identity to *Pax7*). The paired-box sequence of the other zebrafish gene, *pax[zf-b]*, contains only five amino-acid substitutions relative to mouse *Pax2* (ref. 6) and is probably a direct homologue of the murine gene. Transcripts of both genes were first detected in 9–12-h embryos (post-fertilization at 28.5 °C) in the neuroepithelium. Clear-cut expression borders are found at the 16–18-h stage (Figs 2 and 3) when the first axons appear in the brain^{11,12,14}. Later in development, the borders of the two genes in the rostral brain become less pronounced.

Transcripts of *pax[zf-a]* are seen at 18 h in the presumptive thalamic portion of the diencephalon (Figs 2a, e and 3). The posterior border of this expression is located at the forebrain-midbrain junction (as seen in 24-h embryos when morphological borders are visible; Fig. 2b,f). The ventral border coincides with the boundary separating the presumptive thalamic and hypothalamic regions, and anteriorly the expression is limited

‡ To whom correspondence should be addressed.

FIG. 1 *a*, Structure of *pax[zf-a]* cDNAs. The nucleotide and derived amino-acid sequence (single-letter amino-acid notation) of *pax[zf-a]* cDNA is shown starting at the internal *EcoRI* site in the paired-box. The polyadenylation signal is underlined. Base substitutions found in two independent clones are shown above the sequence. The 3' end of the short *pax[zf-a]* clone at position 1,461 is underlined in bold and the location of a 2-base pair (bp) insertion in this clone is indicated (■). *b*, Comparison of the complete paired-box sequence of *pax[zf-b]* and the partial sequence of *pax[zf-a]* to the paired box sequence of *Drosophila prd* (ref. 3) and the murine *Pax1*, *Pax2*, *Pax3*, *Pax7* and *Pax8* genes^{5,6,8-10}. Amino acids identical to those at corresponding positions of the *prd* sequence are indicated by dots. Gaps introduced are represented by dashes. METHODS. Approximately 1.5×10^6 plaques of a 33-h zebrafish *lgt11* cDNA library was screened by plaque hybridization at low stringency²¹ using a mix of the murine *Pax1* *HindIII*-*EcoRI* fragment and a polymerase chain reaction (PCR) product from the zebrafish *Pax1* homologue. Three clones thus isolated represented the *pax[zf-a]* cDNA and a single clone contained the *pax[zf-b]* cDNA.

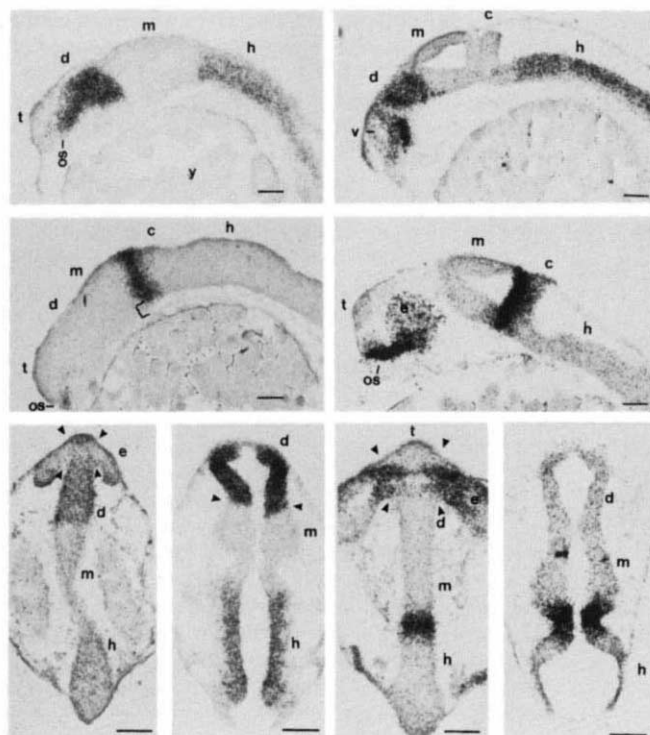
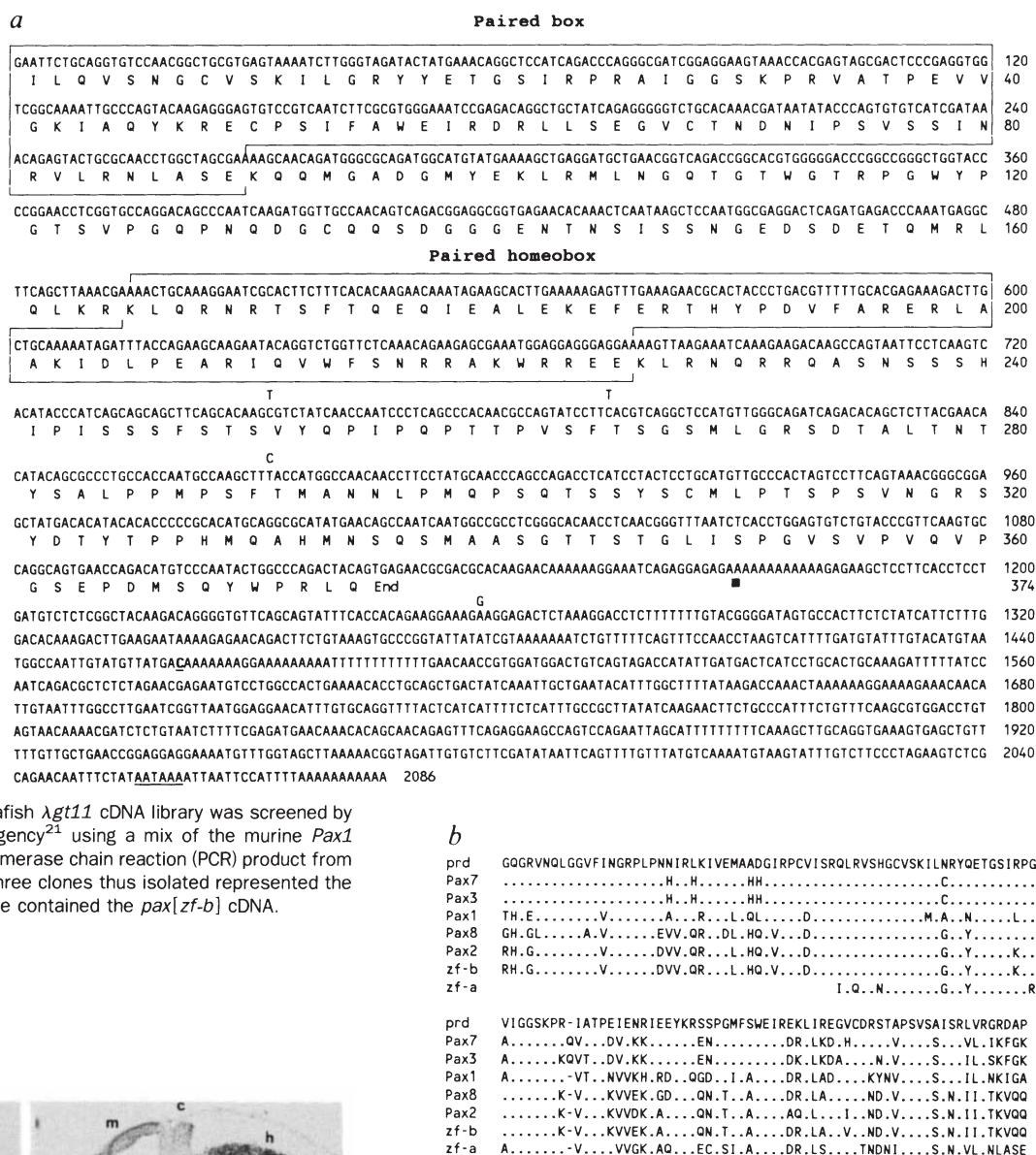


FIG. 2 Spatial localization of transcripts in zebrafish embryos. *a, b*, Sagittal sections of 18-h (*a*) and 24-h (*b*) embryos hybridized with the 600-bp *HindIII*-*EcoRI* fragment from the 3' end of the *pax[zf-a]* cDNA. The position of the ventricle at this stage indicates the diencephalic-telencephalic border. The posterior border of *pax[zf-a]* expression coincides with the diencephalon-midbrain division. The ventral border is located at the division between the thalamic and hypothalamic region of the diencephalon. Owing to a slightly oblique section angle, the presumptive hypothalamus is not seen. *c, d*, Sagittal section of an 18-h (*c*) and a 24-h (*d*) embryo hybridized with the 2.3-kilobase (kb) *EcoRI* insert of the *pax[zf-b]* cDNA. A section from the midline of the embryo shows that the floor plate in the caudal hindbrain does not express the *pax[zf-b]* gene (indicated by a bracket). *e-h*, Transverse sections of 18-h (*e, g*) and 24-h (*f, h*) embryos, hybridized with *pax[zf-a]* (*e, f*) and *pax[zf-b]* (*g, h*) probes. Rostral is at the top. The reported positions of axon tracts (see Fig. 4) are located at landmarks that coincide with the borders of *pax* gene expression as indicated by arrowheads. Note the sharp *pax[zf-a]* expression border between the diencephalon and the midbrain where the tract of the posterior commissure is located (*f*), and the borders of expression at the optic stalk (*e, g*) where the anterior commissure axons and the tract of the postoptic commissure are located. *In situ* hybridization was carried out mainly as described by Molven *et al.*¹⁸. To improve tissue preservation, the acetylation step was not included in the protocol. Abbreviations: c, cerebellum; d, diencephalon; e, eye; h, hindbrain; m, midbrain; os, optic stalk; t, telencephalon; v, ventricle; y, yolk. Scale bars, 50 μ m.

by the optic stalk. Transcripts are also seen in the portions of the optic vesicles overlaying the diencephalon, but no transcripts are detected in the telencephalon. Fewer *pax[zf-a]* transcripts are detected in the ventral portions of the hindbrain (Fig. 2a, b, e, f). The other zebrafish *pax* gene, *pax[zf-b]*, is transcribed in two narrow regions in the rostral brain of the 16–18-h embryo (Figs 2c, g and 3). One region is located at the posterior midbrain. The posterior limit of this region coincides with the furrow separating the mid- and hindbrain (Fig. 2c, d, h). Similar to *Wnt-1*, but in contrast to the *engrailed*-like genes, the expression does not extend towards the cerebellum^{15–19}. Ventrally, *pax[zf-b]* is expressed up to the floor plate which does not express the gene (ref. 2; see brackets in Fig. 2c). Anteriorly, the *pax[zf-b]* expression in the midbrain coincides with the caudal end of the tectal ventricle as seen in 24-h embryos (Fig. 2d, h). The *pax[zf-b]* gene is also transcribed at the base of each optic stalk (Fig. 2c, d, g). In addition, *pax[zf-b]* transcripts can

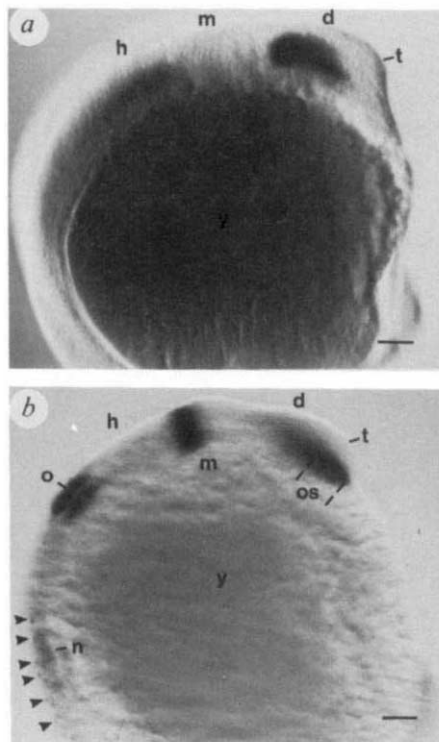


FIG. 3 *In situ* hybridization with digoxigenin-labelled probes (fragments as in Fig. 2). Expression of *a*, *pax[zf-a]* and *b*, *pax[zf-b]* seen on whole mount embryos at 15 h development. The embryos are oriented with the anterior end to the right. Abbreviations; d, diencephalon; h, hindbrain; m, midbrain; n, nephritic primordium; o, otic vesicle; os, optic stalk; t, telencephalon; y, yolk. The arrowheads indicate the positions or single neurons staining for *pax[zf-b]*. Scale bars, 50 μ m.

METHODS. Embryos were fixed for 1 h at room temperature in 4% paraformaldehyde, 4% sucrose, 0.12 mM CaCl_2 , 0.1 M phosphate, pH 7.3. They were then washed for 4 $\times$ 15 min in PBT (1 $\times$ PBS, 0.1% Tween-20, 0.2% BSA), for 10 s in H_2O , 7 min in acetone at -20°C and again for 2 $\times$ 15 min in PBT. PBT was changed to hybridization buffer (50% formamide, 5 $\times$ SSC, 100 μg salmon sperm DNA, 50 μg ml^{-1} heparin and 0.1% Tween-20) and embryos were prehybridized for 1 h at 45°C . The digoxigenin-labelled DNA probe (5 μg ml^{-1} in hybridization buffer) was added and embryos were hybridized overnight at 45°C . Embryos were then washed for 60 min at 45°C in fresh hybridization buffer, for 15 min in 1:1 hybridization buffer-PBT, and twice for 15 min at room temperature in PBT. After the washing steps, 50% of the embryos were used for a 1 h pretreatment of the 1:2,000 diluted anti-digoxigenin antibodies (BMB) in PBT. Pretreated antibodies were then added to the remaining embryos and left at 4°C overnight. Embryos were then washed for 4 $\times$ 15 min in PBT and the staining was performed as recommended in ref. 22. Finally, embryos were dehydrated in ethanol, put for 1 h in methyl salicylate and mounted in Permount.

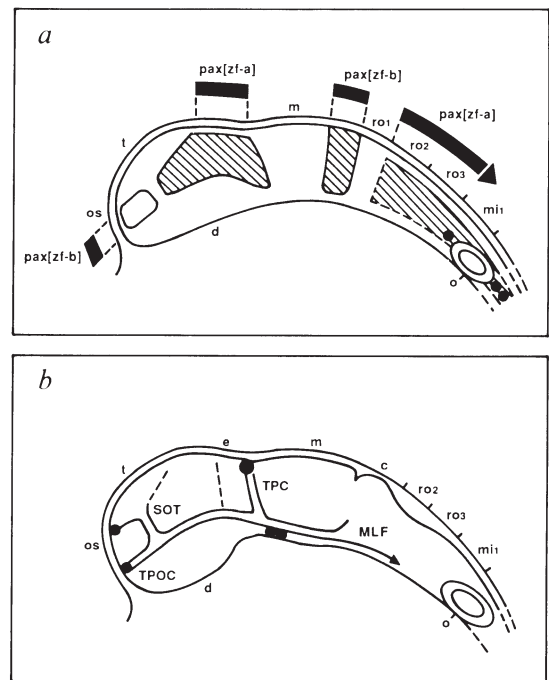


FIG. 4 Comparison of expression patterns and the organization of tracts in the initial axonal scaffold of the rostral brain of zebrafish embryos. *a*, Schematic drawing of the spatial distribution of *pax[zf-a]* and *pax[zf-b]* transcripts in the anterior region of an 18-h embryo. The hatched regions indicate the areas of expression at the surface of the developing brain. An unhatched area at the rostral end shows the location of *pax[zf-b]* transcripts in the base of the optic stalk. The dots mark the isolated neurons expressing *pax[zf-b]* in the hindbrain. *b*, Schematic drawing of the axon tracts present at the surface of the developing brain of 1-day zebrafish embryos. This illustration is based on the semidiagrammatic drawings which have been reported for the 1-day embryo by Wilson *et al.* (ref. 11; S. Wilson, personal communication). The vertical stippled line marks the dorsoventral diencephalic tract which crosses the presumptive thalamus. Abbreviations and symbols: c, cerebellum; d, diencephalon; e, epiphysis; m, midbrain; MLF, medial longitudinal fascicle; o, otic vesicle; os, optic stalk; SOT, supraoptic tract; TPC, tract of the posterior commissure; TPOC, tract of the postoptic commissure. Commissures are represented as filled circles and the rhombomeres of the hindbrain are indicated (ro2, ro3, ml1).

be detected in other tissues, including the otic vesicle, the nephritic primordia and a row of isolated cells along the spinal cord and hindbrain, which according to their position and time of appearance, might be interneurons (Fig. 3b). Expression of the murine *Pax2* gene has not been demonstrated either in the midbrain or in isolated cells in the spinal cord⁷.

The borders of expression of the two zebrafish *pax* genes coincide with pathways of primary axons (Fig. 4). In zebrafish the initial axonal scaffold in the central nervous system (CNS) is rather simple and can easily be correlated with morphological structures^{11,12,14}. The first axon tracts are generated at 17–18 h when the tract of the postoptic commissure divides the diencephalon into the thalamic and hypothalamic region. Slightly later in development the tract of the posterior commissure appears at the junction between the diencephalon and the midbrain^{11,12}. As seen from Figs 2 and 4, the reported positions of both tracts are located at the ventral and caudal expression borders of *pax[zf-a]*, respectively. In addition, Fig. 2e–g demonstrates that the position of the supraoptic tract is located at the junction between the *pax[zf-a]*- and *pax[zf-b]*-expressing regions, whereas the anterior commissure axons are located at the border between the *pax[zf-b]* expressing optic stalk and the forebrain. By contrast, the dorso-ventral diencephalic tract, which starts to form between 18 and 19 h (ref. 23), crosses the

region expressing *pax[zf-a]* (Fig. 4). In the midbrain the medial longitudinal fascicle, descending from the rostral brain to the hindbrain, is in close association with the floor plate and therefore outside the region expressing *pax[zf-b]* (Fig. 2c).

Another aspect of the expression patterns reported here relates to the formation of basic subdivisions of the vertebrate brain^{1,15,19,20} which might be detected at the level of the developmental regulatory genes. The time of appearance, the patterns of expression and the structural homologies of the reported genes would be consistent with these genes having a role in controlling early regionalization in the rostral CNS. As the expression of these *pax* genes precedes axonal outgrowth, one consequence could be the formation of axon pathways at the borders of the *pax*-expressing regions. □

Received 15 May; accepted 3 July 1991.

1. Keynes, R. & Lumsden, A. *Neuron* **2**, 1–9 (1990).
2. Trevarrow, B., Marks, D. L. & Kimmel, C. B. *Neuron* **4**, 669–679 (1990).
3. Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. & Noll, M. *Cell* **47**, 1033–1040 (1986).
4. Bopp, D., Jamet, E., Baumgartner, S., Burri, M. & Noll, M. *EMBO J.* **8**, 1183–1190 (1989).
5. Deutsch, U., Dressler, G. R. & Gruss, P. *Cell* **53**, 617–625 (1988).
6. Dressler, G. R., Deutsch, U., Chowdhury, K., Nornes, H. O. & Gruss, P. *Development* **109**, 787–795 (1990).
7. Nornes, H. O., Dressler, G. R., Knapik, E. W., Deutsch, U. & Gruss, P. *Development* **109**, 797–809 (1990).
8. Plachov, D. *et al.* *Development* **110**, 643–651 (1990).
9. Jostes, B., Walther, C. & Gruss, P. *Mech. Dev.* **33**, 27–38 (1991).
10. Goulding, M. D., Chalepakidis, G., Deutsch, U., Erselius, J. R. & Gruss, P. *EMBO J.* **10**, 1135–1147 (1991).
11. Wilson, S. W., Ross, L. S., Parrett, T. & Easter, S. S. *Development* **108**, 121–145 (1990).
12. Chitnis, A. B. & Kuwada, J. Y. *J. Neurosci.* **10**, 1892–1905 (1990).
13. Treisman, J., Harris, E. & Desplan, C. *Genes Dev.* **5**, 594–604 (1991).
14. Metcalfe, W. K., Myers, P. Z., Trevarrow, B., Bass, M. B. & Kimmel, C. B. *Development* **110**, 491–504 (1990).
15. Wilkinson, D. G., Balles, J. A. & McMahon, A. P. *Cell* **50**, 79–88 (1987).
16. Davis, C. A. & Joyner, A. L. *Genes Dev.* **2**, 1736–1744 (1988).
17. Hatta, K., BreMiller, R., Westerfield, M. & Kimmel, C. B. *Development* **112**, 821–832 (1991).
18. Molven, A., Njåstad, P. R. & Fjose, A. *EMBO J.* **10**, 799–807 (1991).
19. Wilkinson, D. G., Bhatt, S., Cook, M., Boncinelli, E. & Krumlauf, R. *Nature* **341**, 405–409 (1989).
20. Bergquist, H. *Acta zool.* **33**, 117–187 (1952).
21. McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. *Cell* **37**, 403–408 (1984).
22. Tautz, D. & Pfeifle, C. *Chromosoma* **98**, 81–85 (1989).
23. Wilson, S. W. & Easter, S. S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2293–2296 (1991).

ACKNOWLEDGEMENTS. We thank P. Gruss for the *Pax1* probe, K. Zinn for the zebrafish embryonic cDNA library, I. S. Korzh and T. Holm for technical assistance, and S. Wilson, C. Kimmel, M. Westerfield, T. Jessell, N. Holder, P. Gruss and J. Kuwada for critical reading of the manuscript. This work was supported by grants from the Norwegian Research Council NAVF, Norwegian Cancer Society, Aakre Foundation, Hasselblad Foundation and Nansen Foundation.

Hsp104 is a highly conserved protein with two essential nucleotide-binding sites

Dawn A. Parsell, Yolanda Sanchez, Joel D. Stitzel & Susan Lindquist*

Howard Hughes Medical Institute and Department of Molecular Genetics and Cell Biology, 5841 S. Maryland Avenue, University of Chicago, Chicago, Illinois 60637, USA

MOST eukaryotic cells produce proteins with relative molecular masses in the range of 100,000 to 110,000 after exposure to high temperatures¹. These proteins have been studied only in yeast and mammalian cells. In *Saccharomyces cerevisiae*, heat-shock protein hsp104 is vital for tolerance to heat, ethanol and other stresses (ref. 2, and Y.S. *et al.*, manuscript submitted). The mammalian hsp110 protein is nucleolar and redistributes with growth state, nutritional conditions and heat shock^{3–5}. The relationships between hsp110, hsp104 and the high molecular mass heat-shock proteins of other organisms were unknown. We report here that hsp104 is a member of the highly conserved ClpA/ClpB protein family first identified in *Escherichia coli*⁶ and that additional heat-inducible members of this family are present in *Schizosaccharomyces pombe*

and in mammals. Mutagenesis of two putative nucleotide-binding sites in hsp104 indicates that both are essential for function in thermotolerance.

The amino-acid sequence of the hsp104 protein was deduced from the sequence of the *HSP104* gene (Fig. 1a). The sequence predicts a protein of 908 amino acids with a relative molecular mass (M_r) of 102,100 (102.1K), in good agreement with estimates of 104K based on electrophoretic mobility². Two putative

a

```

MNDQCFITERALTILTAAKLASDHQHPQLQPIHILAAFIETPEDGSPVYLQNLIEKGRYDYLKFKVNRNL73
VRIPQQQPAFAETPSYALGKVLQDAAKIQKQKQDSFIAQDHILFALFNDSIIQIFKEAQVDIEIAKQQALE
LRGNTRIDSRGADTNTPLEYLSKYAIDMTQEARQKGLDPVIGREEEIRSTIRVLARIKSNPCLIGEPGIGKI219
ALIEGVAQRIIDDDVPIILQGAFLSLDLAALTAGAKYKGFEEPRKGLKEIEBSKTLIVLEFIDEIHMMLGN
SKDAAAILKPAISRGQLKVGATTNNEYSIVEKDGAFERFQKIEVAEPVSRQTVAILRGLQPKYIEHHGV365
KILDSALVTAQAQAKRYLFPYRRLPSALDLVDISCAGVAVARDSPKEELDSQSIADSSRDKSSRREVECR
IETKRLKLARKEASLQEELEPLRQRYNEEKHGHEELTQAKKLDLENKALVAERRYDTRTAADLRYFAIP511
LIRKQIEKLEDQVAEEERRAGANSIMQNVSDTITSETAARLTGIPVKKLSSENEKLIHMERDLSSEVVGQM
DAIKAVSHAVRLSRGLANRPQASFLFLGLSSGKTELAKKVAGFLFNDEDMIRVDCSELSEKYAVSKLLG657
TTAGYVGYDEGGFLTNQLQYKPYSVLLFDEVEKAHPDVLTVMLQMLDDGRITSGQKKTIDCSNCIVMTSLNG
AEFINSQGSQSKIQUESTKNLVMGAVRQHFRPEFPINRISIVIFNKLRSKAIKHIVDIRLKEIEERFQNDKHY803
INLTQEAQDFLAKYGYSDDMGARPLNRLIQNELINKLALRLKNEIKDKETVNVVLKKGKSRDENVPPEAECC
LEVLPHNEATTGADTLGDDNEDSMIEDDL908

```

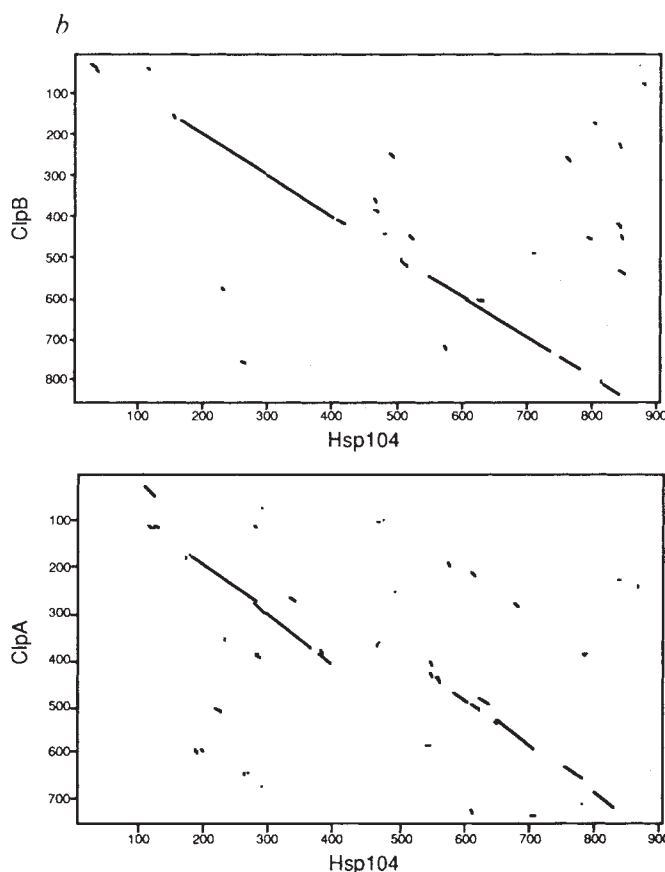


FIG. 1 a, Amino-acid sequence of the *S. cerevisiae* hsp104 protein deduced from the nucleotide sequence of the gene. A deletion series from each end of the *HSP104* gene was generated by digestion of plasmid pYS104 (ref. 2) with exonuclease III. Single-stranded M13 DNAs from both strands of these derivatives were sequenced using a Sequenase kit according to the manufacturer's protocol (US Biochemical). Two regions of the protein show homology to a consensus nucleotide-binding site sequence. Each of these regions is composed of two conserved sequence motifs which are underlined. b, The hsp104 protein sequence was compared with the *E. coli* ClpA and ClpB protein sequences using the MacVector program matrix function (IBI) with a window of 15 amino acids, a minimum score of 40%, a hash value of 2 and the pam250S scoring matrix. The matrices indicate conserved positions.

* To whom correspondence should be addressed.

nucleotide-binding sites were identified in hsp104 by homology with a nucleotide-binding site consensus sequence⁷. The first resembles the catalytic ATP-binding site in the β subunit of F_1 -ATPase; the second is similar to the ATP-binding site of myosin (Fig. 3a)⁸.

A search of GenBank revealed that hsp104 is highly homologous to proteins in the ClpA/ClpB family⁶. These proteins share large (>200 amino acids) blocks of similarity centred around two nucleotide-binding domains. The only member of this group whose function has been characterized is the *E. coli* ClpA protein, which encodes the regulatory subunit of the ATP-dependent Clp (Ti) protease⁹⁻¹³. The *clp* family also includes another *E. coli* gene, *clpB*, three previously cloned genes from other prokaryotes, one from a trypanosome and two from tomato that appear to encode chloroplast proteins⁶. Hsp104 is most similar to the ClpB protein. Matrix plots comparing ClpA and ClpB with hsp104 are shown in Fig. 1b. Hsp104 and ClpB are colinear and homologous throughout. The similarity between hsp104 and ClpA is less dramatic, although marked homology extends over the two large regions encompassing the nucleotide-binding sites.

In light of its sequence similarity to *HSP104*, we investigated the heat inducibility of the *clpB* gene. The *clpB* message is strongly induced at 42 °C (Fig. 2a). Moreover, antiserum raised against a peptide corresponding to hsp104 residues 209–224 (antiserum 2–3) reacts with two heat-inducible *E. coli* proteins whose M_r s are similar to those of the two proteins made from the *clpB* messenger RNA (Fig. 2b)¹⁴. These species are not

observed with lysates from a *clpB* mutant (data not shown). Contemporaneously with our studies, C. Squires, T. Yura and coworkers observed the heat-inducibility of *clpB*^{14,15}. Squires *et al.* also find that *clpB* mutants are more sensitive to high temperatures than wild-type cells¹⁴.

To determine whether heat-inducible homologues of *S. cerevisiae HSP104* exist in other eukaryotic cells, we hybridized portions of the *HSP104* gene at low stringency with total cellular RNAs from different widely divergent species. The amino-terminal half of the *S. cerevisiae* gene hybridized to *S. pombe* and HeLa heat-inducible RNAs of the size that would be expected to encode proteins of 100–110K (Fig. 2c). The same bands were observed when the carboxy-terminal half of the gene was used as a probe (data not shown). These results indicate that *S. pombe* and human cells contain highly conserved heat-inducible genes related to the *S. cerevisiae HSP104* gene.

S. pombe also produces a heat-inducible protein of ~105K which reacts with both antiserum 2–3 and another antiserum (8–1) raised against the C-terminal 15 residues of hsp104 (Fig. 2d). Induction of the 105K protein is stronger with 2–3 than with 8–1. Perhaps the protein is modified on heat shock in a way that affects antiserum recognition, or perhaps the antisera recognize two distinct high- M_r heat-shock proteins that each share epitopes with hsp104.

Antiserum 8–1 also recognized two heat-inducible 100–110K proteins in HeLa and Chinese hamster ovary (CHO) cells (Fig. 2e, f). These proteins correspond precisely to those recognized by an antiserum raised against mammalian hsp110 (Fig. 2f),

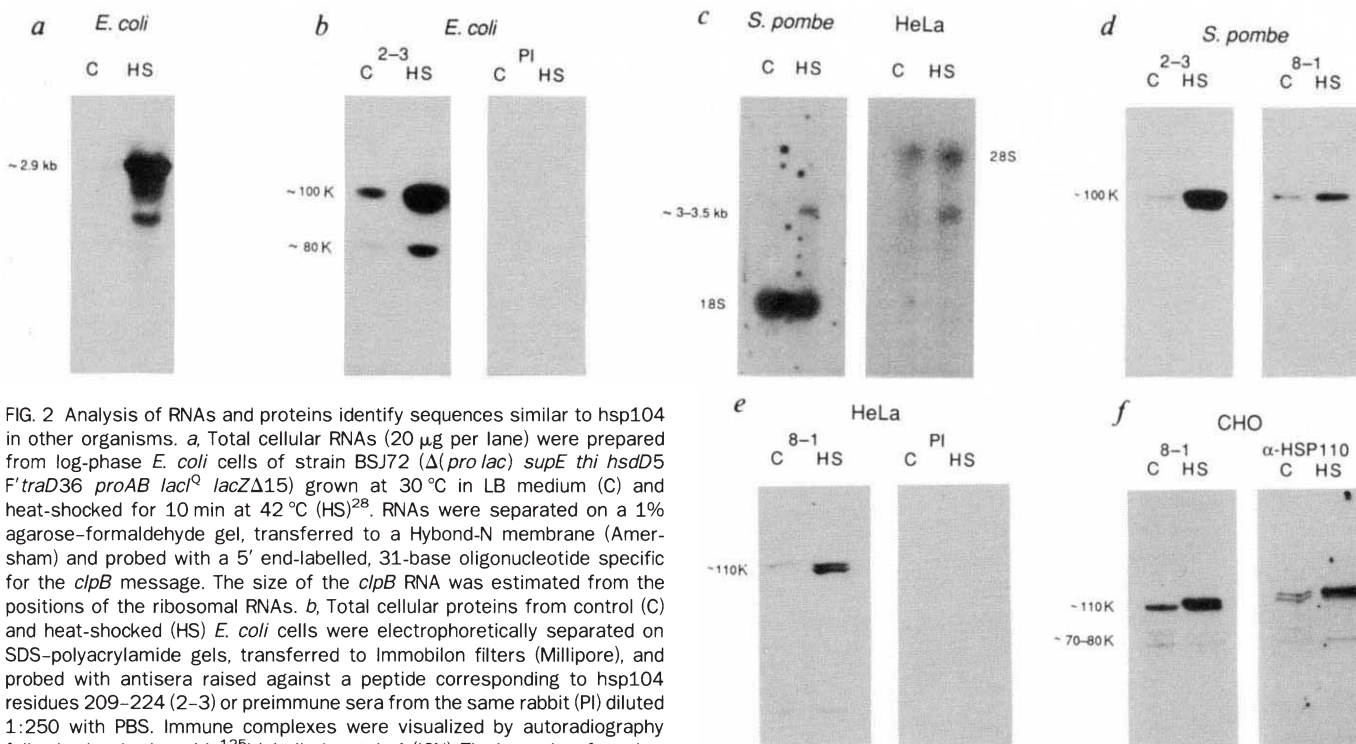


FIG. 2 Analysis of RNAs and proteins identifying sequences similar to hsp104 in other organisms. *a*, Total cellular RNAs (20 µg per lane) were prepared from log-phase *E. coli* cells of strain BSJ72 ($\Delta(pro lac)$ *supE thi hsdD5 F' traD36 proAB lacI^q lacZ Δ 15*) grown at 30 °C in LB medium (C) and heat-shocked for 10 min at 42 °C (HS)²⁸. RNAs were separated on a 1% agarose-formaldehyde gel, transferred to a Hybond-N membrane (Amersham) and probed with a 5' end-labelled, 31-base oligonucleotide specific for the *clpB* message. The size of the *clpB* RNA was estimated from the positions of the ribosomal RNAs. *b*, Total cellular proteins from control (C) and heat-shocked (HS) *E. coli* cells were electrophoretically separated on SDS-polyacrylamide gels, transferred to Immobilon filters (Millipore), and probed with antisera raised against a peptide corresponding to hsp104 residues 209–224 (2–3) or preimmune sera from the same rabbit (PI) diluted 1:250 with PBS. Immune complexes were visualized by autoradiography following incubation with ¹²⁵I-labelled protein A (ICN). The intensity of another band that reacted specifically with the immune serum (not visible in the photograph of Fig. 3b) did not change upon heat shock; it also confirmed equal sample loading on the gel. Molecular weights of the cross-reacting proteins were estimated by comparison with pre-stained molecular weight markers (Bio-Rad). *c*, Total cellular RNAs (10 µg per lane) were prepared²⁹ from log-phase *S. pombe* cells of strain SP223 (*leu1-32, ade6-216, ura4, h⁻*) grown at 28 °C in YPDA medium² (C) and log-phase cells subjected to a 2-h heat shock at 41.5 °C (HS). Total cellular RNAs (5 µg per lane) from control (C) and heat-shocked (HS) HeLa cells were provided by R. Morimoto. Northern analysis was performed as described in *a* using a 1.1-kb fragment from *HSP104* (codons 75–452) labelled with ³²P by random priming as a probe. Hybridization to *S. pombe* and HeLa RNA was carried out at 45 °C and 40 °C, respectively. Diffuse bands of hybridization at the positions of the small *S. pombe* RNA (18S) and the large HeLa RNA (28S) did not change

with heat shock, providing a check of equal sample loading. *d*, Total cellular proteins from control and heat-shocked *S. pombe* cultures, grown as described in *c*, were separated by SDS-PAGE, transferred to Immobilon filters, and probed with antiserum 2–3 or antiserum raised against a peptide corresponding to hsp104 residues 894–908 (antiserum 8–1). *e*, Protein lysates from HeLa cells grown at 37 °C to 80% confluency and maintained at 37 °C (C) or treated at 45 °C for 15 min followed by 6 h of recovery at 37 °C (HS) were analysed as already described with antisera 8–1 or preimmune sera (PI) diluted 1:100 with PBS. *f*, Total cellular proteins from Chinese hamster ovary cells grown at 37 °C and maintained at 37 °C (C) or treated at 45 °C for 15 min followed by 24 h of recovery at 37 °C (HS) were analysed with antiserum 8–1 as described in *e* and with anti-hsp110 antisera diluted 1:400 with PBS.

a

SNPCLIGPGIGKTAIEGVAQ
 NNPLVIGEPGIGKTAIEGVAQ
 NNPLLVGPGSGMGKTAIEGLAW
 GKIGLFGGAGMGKTVFIMELIN
 GKVGLFGGAGMGKTVNMMELIR
 QRELIIQDRGTGKTAIAIDAI
 KIIFVVGSPGSGMGKTAIEGVAQ

Hsp104, residues 206–227
E. coli ClpB, residues 200–221
E. coli ClpA, residues 137–158
 Bovine ATPase β , residues 151–172
E. coli ATPase β , residues 144–165
E. coli ATPase α , residues 163–184
 Adenylate kinase, residues 9–30

b

SNPCLIGEPGIGKTAIEGVAQ
 ↓ V T
 T R
 ASFLFLGLSGSGKTELAKKVAG
 ↓ V T

Hsp104, residues 206–227
 Hsp104, residues 608–629

ASFLFLGLSGSGKTELAKKVAG
 GSFLFLGLPTGVGKTELAKALAN
 GSFLFAGPTGVGKTEVTVOLSK
 QSLILITGESGAGKTEVTKKVIQ
 QSLILITGESGAGKTVNTRKRVIQ
 QSLILITGESGAGKTEVTKKVIC

Hsp104, residues 608–629
E. coli ClpB, residues 597–618
E. coli ClpA, residues 417–438
Dictyostelium myosin, residues 173–194
 Rabbit myosin, residues 172–193
 Nematode myosin, residues 163–184

FIG. 3 **a**, Alignment of two putative nucleotide-binding sequences of hsp104 with nucleotide-binding domains of other proteins. Identical amino acids are boxed. References: Hsp104, this work; ClpA and ClpB, ref. 6; bovine and *E.*

coli ATPase subunits, rabbit and nematode myosins, ref. 8; *Dictyostelium* myosin, ref. 30. **b**, Site-directed mutagenesis of HSP104. Synthetic oligonucleotides were used to direct the desired mutations to the HSP104 gene using a Muta-Gene *in vitro* mutagenesis kit (Bio-Rad). Single-stranded plasmid DNAs were sequenced to identify the mutations and confirm their fidelity. Amino-acid substitutions that destroy the function of hsp104 are indicated by downward arrows; those that have little or no effect on thermotolerance are indicated by upward arrows.

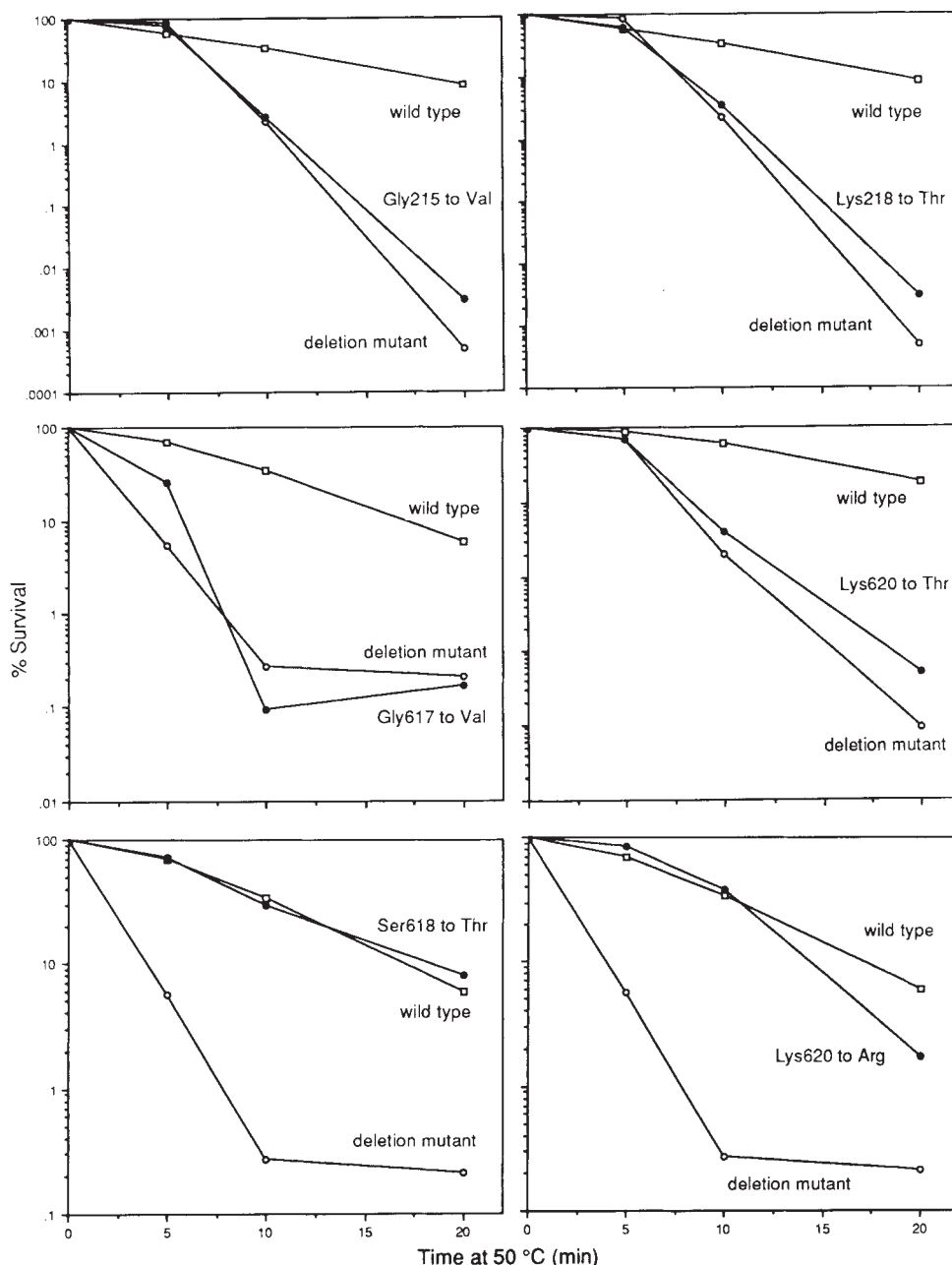


FIG. 4 Mutations in the two putative nucleotide-binding sequences of hsp104 affect induced thermotolerance. Early log-phase *S. cerevisiae* cells (2×10^6 cells per ml) deleted for the chromosomal HSP104 gene (Δ 104) and carrying a centromere-containing plasmid (pYS104) encoding a wild-type ($\square$) or mutant ($\bullet$) HSP104 gene or carrying the vector alone ($\circ$) were grown at 25 °C in minimal medium appropriately substituted to maintain the plasmid². Cultures were preincubated at 37 °C for 30 min before transfer to 50 °C for the times indicated. After treatment at 50 °C, cells were placed on ice, diluted with ice-cold YPDA and plated on YPDA.

suggesting that hsp110 may be the product of the heat-inducible mammalian transcript recognized by the *HSP104* probes. Although C termini of previously identified *clp*-related genes are not homologous⁶, precedent exists for the conservation of C termini among the eukaryotic members of an hsp family which does not extend to its prokaryotic and organellar members. In fact, the C terminus of hsp104 (AspAspLeuAsp) resembles the sequence GluGluValAsp, which occurs at the C terminus of virtually all eukaryotic hsp70 and hsp90 proteins¹. Notably, eight of the nine *clp*-like genes already described are prokaryotic or encode proteins containing signal sequences that should target them to organelles⁶. The only eukaryotic *clp* gene that does not contain a signal sequence is from trypanosomes and encodes a protein with a C-terminal sequence quite similar to that of hsp104 (AspGluTrpGlu). Perhaps, like hsp70s and hsp90s, eukaryotic non-organellar hsp100 proteins share conserved negatively charged C-terminal sequences.

Hsp104 contains two putative nucleotide-binding sites (Figs 1a, 3a)⁷. Both sites contain two characteristic elements: a GlyX₂GlyXGlyLysThr sequence which forms a glycine-rich flexible loop and, about 60 residues downstream, a stretch of at least four hydrophobic amino acids terminated by an aspartate^{8,16}. X-ray crystallographic and NMR studies of nucleotide-binding pockets indicate that the glycine residues in the first element maintain the loop conformation necessary for proper nucleotide binding and that the lysine residue interacts with one of the phosphoryl groups of the bound nucleotide^{8,16}. ClpA contains similar sequences and is known to bind and hydrolyse ATP *in vitro*^{9,10}. To determine whether these sites are essential for the function of hsp104, we introduced site-directed mutations into the *HSP104* gene that resulted in separate single amino-acid substitutions in these regions. Based upon mutational analysis of other nucleotide-binding pockets^{17,18}, we designed four substitutions (downward-pointing arrows, Fig. 3b) to inactivate the sites without disrupting the overall structural stability of the protein. The second glycine in each of the loops was changed to valine and, separately, the lysine to threonine (Fig. 3b). The ability of hsp104 to protect the cell from extreme heat was destroyed by each of these substitutions (Fig. 4, top and centre panels). In the second potential nucleotide-binding site we also constructed two single amino-acid substitutions (upward-pointing arrows, Fig. 3b) that we expected to be less damaging. The non-conserved serine was replaced with threonine and, separately, the conserved lysine was replaced with another positively charged residue, arginine. Cells expressing these variants showed almost wild-type thermotolerance (Fig. 4, bottom panel). None of the substitutions affected the amount of hsp104 made or its mobility on SDS-polyacrylamide gels (data not shown), suggesting that they do not affect the overall structure of the protein or increase its sensitivity to intracellular proteolysis. These preliminary results indicate that hsp104 contains two nucleotide-binding sites and that both are required for its function in thermotolerance.

How might hsp104 protect cells from extreme stress? One possibility, suggested by sequence homology to ClpA, is that proteins damaged by high temperature are toxic and hsp104 affords thermotolerance by promoting their degradation. Eukaryotic cells contain an ATP-dependent proteolytic complex which degrades ubiquitinated substrates. This complex is composed of a 10–20 subunit ATP-independent particle (the proteasome), which may be homologous to the *E. coli* ClpP particle^{19,20}, and other proteins which include a 100K and a 110K species^{21,22}. The latter might be hsp100s which function like ClpA to make the complex ATP-dependent. But, we have failed to detect a proteolytic defect in a strain lacking hsp104 (D.A.P. and S.L., unpublished observation) and its phenotype is quite different from those of strains lacking proteasome subunits or components of the ubiquitin system^{2,23–26}.

We suggest that rather than operating strictly as proteolytic regulators, hsp100s function more generally to protect cells from

extreme stress by controlling the aggregation or denaturation of vital cellular structures. Such a role is reminiscent of the proposed function of another heat-shock protein, hsp70 (ref. 27). Similarity in the biological effects of hsp100s and hsp70s is likely, given that the thermotolerance defect in *S. cerevisiae* cells lacking hsp104 is largely suppressed by the overexpression of hsp70 (Y.S., E.A. Craig and S.L., manuscript in preparation). It is notable that the proteolytic activity of the proteasome can be activated *in vitro* by mild denaturing treatments (ref. 22 and refs therein). If hsp100 proteins function *in vivo* to activate a proteolytic complex, perhaps they do so through a general disaggregation or denaturation activity. □

Received 10 June; accepted 6 August 1991.

1. Lindquist, S. & Craig, E. A. *Rev. Genet.* **22**, 631–677 (1988).
2. Sanchez, Y. & Lindquist, S. L. *Science* **248**, 1112–1115 (1990).
3. Subjeck, J. R., Shyy, T., Shen, J. & Johnson, R. J. *J. Cell. Biol.* **97**, 1389–1395 (1983).
4. Shyy, T. T., Subjeck, J. R., Heinaman, R. & Anderson, G. *Cancer Res.* **46**, 4738–4745 (1986).
5. Welch, W. J. & Suhan, J. P. *J. Cell Biol.* **101**, 1198–1211 (1985).
6. Gottesman, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 3513–3517 (1990).
7. Brenner, S. *Nature* **334**, 528–530 (1988).
8. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. *EMBO J.* **1**, 945–951 (1982).
9. Katayama-Fujimura, Y., Gottesman, S. & Maurizi, M. R. *J. Biol. Chem.* **262**, 4477–4487 (1987).
10. Hwang, B. J., Park, W. J., Chung, C. H. & Goldberg, A. L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5550–5554 (1987).
11. Katayama, Y. *et al. J. Biol. Chem.* **263**, 15226–15236 (1988).
12. Hwang, B. J., Woo, K. M., Goldberg, A. L. & Chung, C. H. *J. Biol. Chem.* **263**, 8727–8734 (1988).
13. Gottesman, S., Clark, W. P. & Maurizi, M. R. *J. Biol. Chem.* **265**, 7886–7893 (1990).
14. Squires, C. L., Pedersen, S., Ross, B. M. & Squires, C. J. *Bact.* **173**, 4254–4262 (1991).
15. Kitagawa, M., Wada, C., Yoshioka, S. & Yura, T. *J. Bact.* **173**, 4247–4253 (1991).
16. Fry, D. C., Kuby, S. A. & Mildvan, A. S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 907–911 (1986).
17. Liu, Q. & Summers, W. C. *Virology* **163**, 638–642 (1988).
18. Xia, Z. & Storm, D. R. *J. Biol. Chem.* **265**, 6517–6520 (1990).
19. Tanaka, K. *et al. Biochem. biophys. Res. Commun.* **164**, 1253–1261 (1989).
20. Maurizi, M. R. *et al. J. Biol. Chem.* **265**, 12536–12545 (1990).
21. Hough, R., Pratt, G. & Rechsteiner, M. *J. Biol. Chem.* **262**, 8303–8313 (1987).
22. Eytan, E., Ganoh, D., Armon, T. & Herskko, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7751–7755 (1989).
23. Heinemeyer, W., Kleinschmidt, J. A., Saidowsky, J., Escher, C. & Wolf, D. H. *EMBO J.* **10**, 555–562 (1991).
24. Emori, Y. *et al. Molec. cell. Biol.* **11**, 344–353 (1991).
25. Finley, D., Ozkaynak, E. & Varshavsky, A. *Cell* **48**, 1035–1046 (1987).
26. Seufert, W. & Jentsch, S. *EMBO J.* **9**, 543–550 (1990).
27. Pelham, H. R. *Cell* **46**, 959–961 (1986).
28. Gilman, M. in *Current Protocols in Molecular Biology* (ed. Ausubel, F. M.) 441–444 (Greene and Wiley-Interscience, New York, 1987).
29. Vijayraghavan, U., Company, M. & Abelson, J. *Genes Dev.* **3**, 1206–1216 (1989).
30. Warrick, H. M., DeLozanne, A., Leinward, L. A. & Spudis, J. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9433–9437 (1986).

ACKNOWLEDGEMENTS. Antisera against mammalian hsp110 and nitrocellulose filters containing control and heat-shocked CHO lysates were kindly provided by J. Subjeck and J. W. Shen. We thank C. Squires and T. Yura for sharing results before publication, R. Morimoto for control and heat-shocked HeLa RNA samples, J. Vogel for the *S. pombe* experiment shown in Fig. 2c, R. Petersen for assistance with HeLa lysates, D. Kang for photography, and L. Arwood, J. Vogel, R. Petersen and J. Feder for comments on the manuscript. This work was supported by the National Institutes of Health, the Howard Hughes Medical Institute, the Jane Coffin Childs Memorial Foundation for Medical Research and the Juan March Foundation. The GenBank accession number of *HSP104* is M67479.

Recognition of UGA as a selenocysteine codon in Type I deiodinase requires sequences in the 3' untranslated region

Marla J. Berry, Laila Banu, Yoyi Chen, Susan J. Mandel, J. David Kieffer, John W. Harney & P. Reed Larsen

Howard Hughes Medical Institute Laboratory and Thyroid Division, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 75 Francis Street, Boston, Massachusetts 02115, USA

SELENOCYSTEINE is incorporated cotranslationally at UGA codons, normally read as stop codons, in several bacterial proteins^{1,2} and in the mammalian proteins glutathione peroxidase (GPX)^{3–5}, selenoprotein P⁶ and Type I iodothyronine 5' deiodinase (5'DI)⁷. Previous analyses in bacteria have suggested that a stem-loop structure involving the UGA codon and adjacent sequences is necessary and sufficient for selenocysteine incorporation into

formate dehydrogenase and glycine reductase^{2,8,9}. We used the recently cloned 5'DI to investigate selenoprotein synthesis in eukaryotes. We show that successful incorporation of selenocysteine into this enzyme requires a specific 3' untranslated (3'ut) segment of about 200 nucleotides, which is found in both rat and human 5'DI messenger RNAs. These sequences are not required for expression of a cysteine-mutant deiodinase. Although there is little primary sequence similarity between the 3'ut regions of these mRNAs and those encoding GPX, the 3'ut sequences of rat GPX can substitute for the 5'DI sequences in directing selenocysteine insertion. Computer analyses predict similar stem-loop structures in the 3'ut regions of the 5'DI and GPX mRNAs. Limited mutations in these structures reduce or eliminate their capacity to permit 5'DI translation. These results identify a 'selenocysteine-insertion sequence' motif in the 3'ut region of these mRNAs that is essential for successful translation of 5'DI, presumably GPX, and possibly other eukaryotic selenocysteine-containing proteins.

Type I iodothyronine deiodinase catalyses the first step in thyroid hormone action, the monodeiodination of the prohormone, thyroxine (T₄), to form the active thyroid hormone, 3,5,3'-triiodothyronine (T₃). Selenocysteine is required for normal 5' deiodinase activity in the rat enzyme⁷. A cysteine mutant is also functional, albeit with a 10-fold higher apparent Michaelis constant (K_m) for the preferred substrate, 3,3',5'-triiodo-

thyronine (reverse T₃)^{7,10}. The open reading frame of the 2.1 kilobase (kb) rat 5'DI mRNA begins at nucleotide 7 and ends at 780, and the UGA (selenocysteine) codon is located at nucleotides 382–384 (ref. 7, Fig. 1a). Deletion of sequences 3' to nucleotide 907 in the wild-type 5'DI complementary DNA resulted in complete loss of deiodinase activity in either JEG-3 or COS-7 cell transient transfections (Fig. 1a). But deletion of the same sequences had no effect on the activity expressed by the cysteine mutant. Sequences in the 3'ut region are therefore required for selenocysteine, but not cysteine, incorporation. These results argue against a regulatory role for the deleted sequences in transcription, RNA processing, or stability. Further studies were done to identify the specific sequences in the wild-type cDNA necessary for selenocysteine incorporation. Deletion analyses identified sequences between nucleotides 1,440 and 1,615 as being required for expression of 5'DI activity (Fig. 1a). To confirm that these sequences were required only at the level of translation, RNA was prepared by *in vitro* transcription of plasmids containing either the full-length 5'DI cDNA, the cDNA truncated at nucleotide 1,580 or the cDNA lacking nucleotides 1,278–1,495. The relative 5'DI activity produced by injection of these RNAs into *Xenopus* oocytes paralleled that produced by transfection, evidence that impaired translation causes the reduced expression by the 3'ut mutants.

FIG. 1 a, Deletion and inversion mutations of rat 5'DI cDNA 3'ut region. Wild-type and mutant rat 5'DI constructs were assayed for production of 5'DI activity after transient transfection in JEG-3 or COS-7 cells. Deiodinase activity at the level of the wild-type rat 5'DI construct is defined as 100%, and was equivalent to 5' deiodination of 2 pmol reverse T₃ per min per mg protein for TGA-containing constructs and 1 pmol reverse T₃ per min per mg protein for TGT-containing constructs, in JEG cell extracts. b, Rat 5'DI constructs containing 3'ut sequences from rat or human 5'DI or rat GPX cDNAs. Constructs containing either rat or human 5'DI or rat GPX 3'ut sequences adjacent to rat 5'DI coding sequences were assayed for production of 5'DI activity as above.

METHODS. Isolation of the rat 5'DI cDNA has been reported previously⁷. Deletion mutants were constructed in either Bluescript (Stratagene) or CDM-8¹² by digestion with restriction enzymes and conversion to blunt ends with DNA polymerase large fragment when ends were incompatible, followed by religation. Mutants constructed in Bluescript were subsequently cloned into CDM-8 at appropriate sites. Site-directed mutagenesis was done as described previously⁸ using the P-select *in vitro* mutagenesis system (Promega). Using the rat 5'DI cDNA as a probe, we have isolated a partial human 5'DI cDNA from a human liver cDNA library (kindly provided by B. Seed), by standard hybridization screening techniques. The rat GPX cDNA was a kind gift of Ye-Shih Ho¹¹. Chimaeric constructs were generated by PCR-amplification using 5' oligonucleotides encoding an *Xma*I site and 3' oligonucleotides encoding a *Not*I site adjacent to sequences specific for the region to be amplified. PCR products were cloned into *Xma*I + *Not*I cut CDM-8 containing the 5'DI cDNA. This vector fragment contains the 5'DI coding region and 126 bp of 3'ut, and is nonfunctional for deiodinase activity in transient transfection and oocyte injection assays. JEG-3 or COS-7 cells were transfected with calcium phosphate-DNA precipitates as described previously¹³. Transfection efficiencies were monitored by assay of human growth hormone in the media, produced by a cotransfected constitutive thymidine kinase promoter-directed human growth hormone-expressing plasmid¹³. Cell sonicates were assayed for 5' deiodination of [¹²⁵I]-reverse T₃ in reactions containing 10–250 µg protein, 300 nM [¹²⁵I]-reverse T₃, and 10 mM DTT in 0.1 M potassium phosphate, pH 6.9, 1 mM EDTA. Incubations were for 1 h in 500 µl volume.

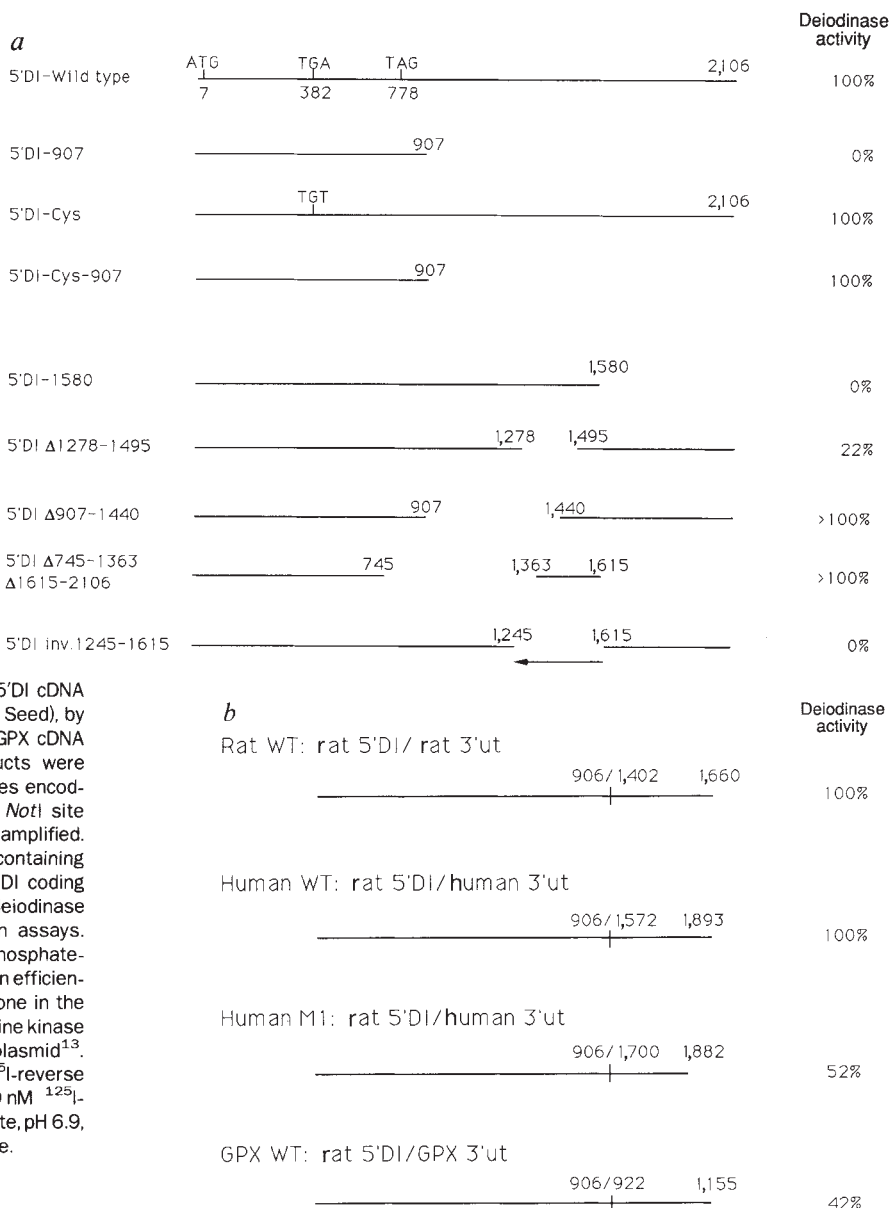
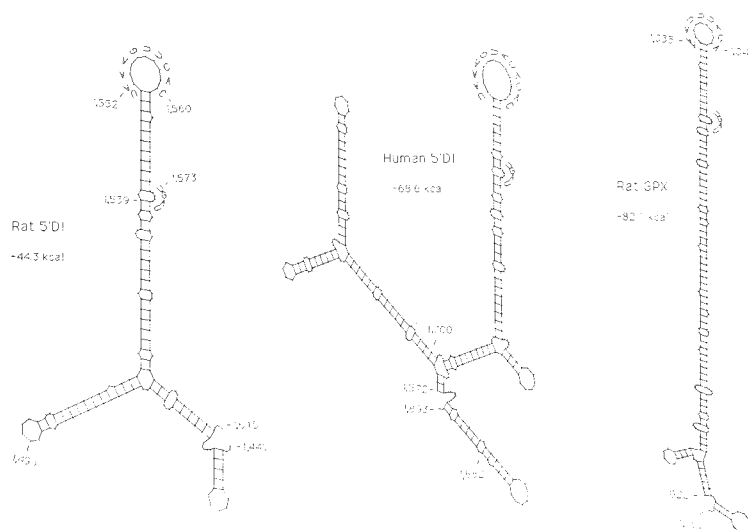


FIG. 2 Predicted secondary structures in the 3'ut regions of selenocysteine-encoding RNAs. Sequences from the 3'ut regions of the rat 5'DI⁸, human 5'DI and rat GPX¹¹ are shown. The positions of deletions which resulted in partial or complete loss of function are indicated. Structure analysis was performed using the FOLD program of the University of Wisconsin Genetics Computer Group software¹⁴.



We compared the 3'ut sequences of the cDNA for the rat 5'DI with the sequence of a cDNA for the human 5'DI recently isolated in our laboratory (manuscript in preparation). Although the 3'ut sequences are about 55% conserved overall, a region of about 79% identity (nucleotides 1,642–1,819) corresponded to the essential 3'ut sequences (1,440–1,615) identified in the rat 5'DI cDNA. A construct in which conserved human 3'ut sequences (1,572–1,893) were inserted downstream of the rat coding region (human WT) produced deiodinase activity equal to that produced by the construct containing the essential rat 3'ut sequences (Fig. 1b). A second construct (human M1) containing an abbreviated 3'ut segment (1,700–1,882) produced $52 \pm 3\%$ of the wild-type amount of activity. Examination of the 3'ut region of the rat GPX mRNA revealed less than 38% primary sequence similarity to the conserved 5'DI sequence. But addition of nucleotides 922–1,155 from the 3'ut of the rat GPX cDNA to the coding region of the rat 5'DI cDNA restored $42 \pm 3\%$ of the 5'DI activity (Fig. 1b). The distance between the UGA codon and the midpoint of the functional 3'ut sequences varies from 548 nucleotides in the rat GPX mRNA¹¹ to 1,145 in the rat⁷ and 1,409 in the human 5'DI mRNAs, respectively. Two constructs with reduced spacing between the coding region and the 3'ut sequences produced levels of deiodinase activity about 30–40% higher than did the wild-type cDNA (Fig. 1a), suggesting that the spacing between the 3' sequences and the UGA codon influences the efficiency of selenocysteine insertion.

The lack of primary sequence similarity between the 5'DI and GPX 3'uts suggested that secondary structures may be involved in regulating translation of the UGA codons in these mRNAs. Analyses of these sequences predicted the stem-loop structures shown in Fig. 2, all of which have high negative free energy. All three contain a large stem-loop with three adjacent As in the loop and an unpaired UGAU in the stem. The 5'DI sequences also predict a smaller putative stem-loop located 5' to the larger one. The smaller stem-loop is disrupted in the rat 5'DI $\Delta 1278$ –1495 construct and deleted in the human M1 construct, both of which have impaired activity. Thus, the smaller loops may contribute to the process of selenocysteine insertion, but are not absolutely essential. To test the function of the large putative stem-loop structures, the deletions shown in Fig. 3 were generated. Removal of either the 35-base pair (bp) stem-loop (rat M1) or the 9-bp loop (rat M2) resulted in loss of the capacity to confer deiodinase expression. Similarly, deletion of the 8-bp GPX loop (GPX M1) inactivated this construct.

Reticulocyte lysates translate rat 5'DI mRNA inefficiently, producing small amounts of full-length protein of relative molecular mass about 27,000 (M_r , 27 K), with most of the

translated product being the roughly 14K protein predicted by termination at the UGA codon⁷. If the 3'ut sequences are involved in selenocysteine codon recognition, the ratio of 27 to 14K protein should be reduced in *in vitro* translations of 3'ut mutant transcripts. We prepared *in vitro* transcripts of wild-type and mutant 5'DI constructs, translated the RNA *in vitro*, immunoprecipitated the translation products with 5'DI specific antisera generated against a rat 5'DI amino-terminal peptide,

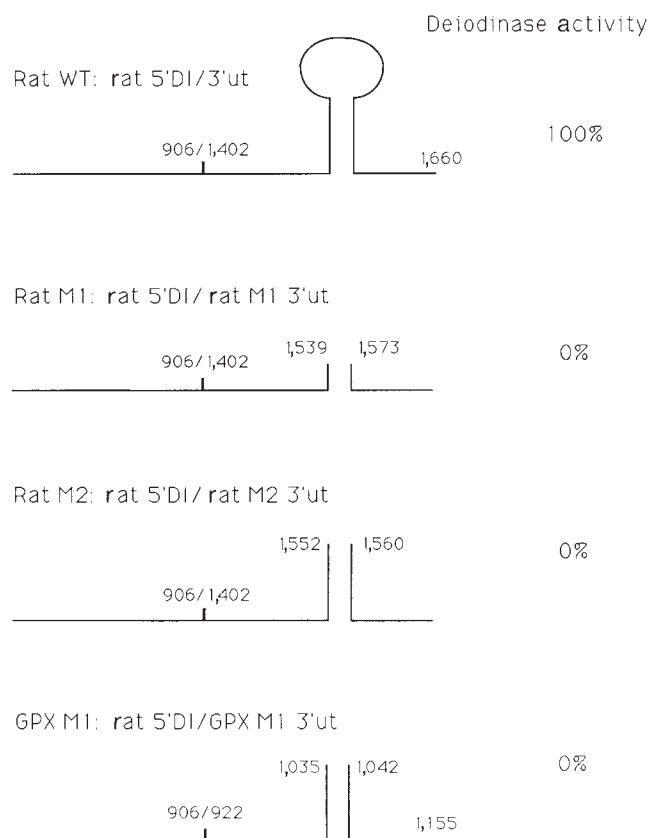


FIG. 3 Deletion mutations in the stem-loop regions of rat 5'DI and GPX mRNAs. PCR deletions were generated as described in *PCR Protocols*¹⁵, and cloned into the vector fragment described in the legend to Fig. 1. Deiodinase assays were done as described in the legend to Fig. 1.

Rat 5'DI (1523)	TTTATATTTG	TTTATGATGG	.TCACAGTGT	AAAGTTCACA	CAGCTGTG.A	CTTGATTTTT	..AAAAATGT	CGGGAAGA
Human 5'DI (1728)	TACATATTTG	TTTATGATGG	.CCACAGCCT	AAAGTACACA	CGGCTGTG.A	CTTGATTTCAA	AAGAAAATGT	TATAAGA.
5'DI Consensus	T..ATATTTG	TTTATGATGG	..CACAG..T	AAAGT.CACA	C.GCTGTG.A	CTTGATT...	...AAAATGT	A...
Rat GPX (1004)	GGGAGGTTTT	TCCATGACGG	TGTTTCCTCT	AAATTTACAT	GGAGAAAC.A	CCTGATTTCC	AGAAAAATCC	CCTCAGAT
Mouse GPX (1168)	GGGCGGTTCT	TCCATGATGG	TGTTTCCTCT	AAATTTGCAC	GGAGAAAC.A	CCTGATTTCC	AGGAAAATCC	CCTCAGAT
Human GPX (942)	GGGGTTTTCA	TCTATGAGGG	TGTTTCCTCT	AAACCTACGA	GGGAGGAACA	CCTGATCTTA	CAGAAAATCC	CACCTCGA
Bovine GPX (895)	AGGGATTTTG	CCCATGAAGG	TGTTTCCTCT	AAACCTACGT	GGAGGAAT.G	CCTGATGTCC	AGGAAAATCC	CCTGAGGT
GPX consensus	.GG...TT..	.C.ATGA.GG	TGTT..CCTCT	AAA..T.C..	GG....A...	CCTGAT.T..	...AAAAT.C	C.....
5'DI consensus	T..ATATTTG	TTTATGATGG	..CACAG..T	AAAGT.CACA	C.GCTGTG.A	CTTGATT...	...AAAATGT	A...
GPX consensus	.GG...TT..	.C.ATGA.GG	TGTT..CCTCT	AAA..T.C..	GG....A...	CCTGAT.T..	...AAAAT.C	C.....
SECIS consensus	TT..	...ATGA.GG	T	AAA.....	A...	C.TGAT....	...AAAAT..	

FIG. 4 Sequence similarities in the stem-loop regions of the rat and human 5'DI, and mammalian GPX cDNAs. Analysis was done using the LINEUP

program¹⁴. Nucleotide numbers are shown in parentheses.

and quantitated the 27 and 14K [³⁵S] methionine-labelled products after SDS-PAGE. The ratio of 27 to 14K protein was 0.17 ± 0.03 for the rat WT, 0.08 ± 0.04 for the human M1, and 0.013 ± 0.004 for the rat M2 construct. These results establish the UGA recognition function of the 3'ut regions.

The 3'uts of the various selenocysteine-encoding mRNAs have little primary sequence similarity. Alignment of the proposed stem-loop regions of the rat and human 5'DI mRNAs with the 3'ut regions of the mammalian GPX mRNAs identifies the conserved nucleotides shown in Fig. 4. Inversion of the sequences between 1,245 and 1,615 in the rat 5'DI cDNA resulted in loss of 5'DI activity (Fig. 1a). The predicted stem-loop in this construct is similar to the wild-type, but as the sequence is complementary, none of the conserved bases are present. The 3'ut of human plasma GPX⁵ contains no significant similarity to the sequences in Fig. 4, however, this sequence is not full length, as it lacks a polyadenylation signal sequence and poly (A) tail. The recently reported cDNA for rat selenoprotein P (ref. 6) contains a 3'ut region of >1,600 nucleotides, and secondary structure analyses predict 15 stem-loops with free energies of -20 kcal or less. Six of these contain UAAA or AAA sequences in the loop. Because of the numerous potential stem-loops in this long 3'ut, the identification of specific regions involved in selenocysteine insertion may require deletion mapping or other functional analyses.

Previous studies of the *Escherichia coli* FDH mRNA established that only the sequences immediately adjacent to the UGA codon are required for its recognition as a selenocysteine codon⁸. We have not examined the role of such sequences in the 5'DI mRNA. Our results establish that a roughly 200 nucleotide segment located more than 1 kb downstream of the UGA in the 5'DI mRNAs is essential for insertion of selenocysteine into this protein. Although most of our results were obtained using transient expression techniques, we have demonstrated that these sequences are required for *in vivo* and *in vitro* translation of the intact, fully functional protein. Thus we may term this segment of these mRNAs a selenocysteine-insertion sequence (SECIS) motif. The requirement for a SECIS motif in the 5'DI mRNAs for successful translation of this protein, and the presence of sequences with similar function in the GPX mRNAs, identifies a previously unrecognized regulatory step in the expression of genes encoding eukaryotic selenocysteine-containing proteins. Such motifs may be required in eukaryotic expression vectors for insertion of the more reactive selenium in place of sulphur in sulphhydryl active-site proteins for purposes of biochemical or structural analyses. How they function remains to be determined. □

Received 30 May; accepted 29 July 1991.

1. Zinoni, F., Birkmann, A., Stadtman, T. & Bock, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4650-4654 (1986).
2. Garcia, G. E. & Stadtman, T. C. *J. Bact.* **174**, 2093-2098 (1991).
3. Chambers, I. et al. *EMBO J.* **5**, 1221-1227 (1986).
4. Mullenbach, G. T. et al. *Protein Eng.* **2**, 239-246 (1988).
5. Takahashi, K. et al. *J. Biochem.* **108**, 145-148 (1990).
6. Hill, K. E., Lloyd, R. S., Yang, J. G., Read, R. & Burk, R. F. *J. Biol. Chem.* **266**, 10050-10053 (1991).
7. Berry, M. J., Banu, L. & Larsen, P. R. *Nature* **349**, 438-440 (1991).
8. Zinoni, F., Heider, J. & Bock, A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4660-4664 (1990).

9. Bock, A. *Molec. Microbiol.* **5**, 515-520 (1991).
10. Berry, M. J., Kieffer, J. D., Harney, J. W. & Larsen, P. R. *J. Biol. Chem.* **266**, 14155-14158 (1991).
11. Ho, Y.-S., Howard, A. J. & Crapo, J. *Nucleic Acids Res.* **16**, 5207 (1988).
12. Aruffo, A. & Seed, B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8573-8577 (1987).
13. Brent, G. A. et al. *Molec. Endocrinol.* **3**, 1996-2004 (1989).
14. Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387-395 (1984).
15. Higuchi, R. in *PCR Protocols* (eds Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J.) 177-183 (Academic, New York, 1990).

ACKNOWLEDGEMENTS. We thank B. Seed and D. D. Moore for critical reading of the manuscript. This work was supported by the NIH.

Structure of human cyclophilin and its binding site for cyclosporin A determined by X-ray crystallography and NMR spectroscopy

Jörg Kallen, Claus Spitzfaden*, Mauro G. M. Zurini, Gerhard Wider*, Hans Widmer, Kurt Wüthrich* & Malcolm D. Walkinshaw†

Preclinical Research, Sandoz Pharma AG, 4002 Basel, Switzerland

* Institut für Molekularbiologie und Biophysik, ETH-Hönggerberg, 8093 Zürich, Switzerland

THE protein cyclophilin is the major intracellular receptor for the immunosuppressive drug cyclosporin A (ref. 1). Cyclosporin A acts as an inhibitor of T-cell activation and can prevent graft rejection in organ and bone marrow transplantation². Cyclophilin may be responsible for mediating this immunosuppressive response. Cyclophilin also catalyses the interconversion of the *cis* and *trans* isomers of the peptidyl-prolyl amide bonds of peptide and protein substrates^{3,4}. Here we report the X-ray crystal structure of human recombinant cyclophilin complexed with a tetrapeptide and the identification, by nuclear magnetic resonance spectroscopy, of the specific binding site for cyclosporin A. Cyclophilin has an eight-stranded antiparallel β -barrel structure. The prolyl isomerase substrate-binding site is coincident with the cyclosporin-binding site. These results may help to provide a structural basis for rationalizing the immunosuppressive function of the cyclosporin-cyclophilin system and will also be important in the design of improved immunosuppressant drugs.

Cyclophilin (relative molecular mass 17,800 (M_r , 17.8K)) consists of a single polypeptide chain with 165 amino-acid residues (Fig. 1a). Cyclosporin A (CsA) is a cyclic undecapeptide with the sequence c-(MeBmt-Abu-Sar-MeLeu-Val-MeLeu-Ala-D-Ala-MeLeu-MeLeu-MeVal), where the prefix Me indicates *N*-methylation and the uncommon amino acids in positions 1 and 2 are (4R)-4-((E)-2-butenyl)-4, *N*-dimethyl-L-

† To whom correspondence should be addressed.

TABLE 1 Crystallographic and multiple isomorphous replacement (MIR) data

Data set	Space group $P2_12_12_1$ ($a=108.2$ Å, $b=123.0$ Å, $c=35.8$ Å)					
	Native (1)	Native (2)	EMTS*	PHMPS	K ₂ PtCl ₄ (1)	K ₂ PtCl ₄ (2)
Number of collected reflections	43,082	42,788	43,308	26,385	23,713	23,600
$R_{\text{sym}}(I)$ in parentheses†	(6.9%)	(7.6%)	(7.2%)	(8.0%)	(10.2%)	(8.7%)
Number of unique reflections‡	11,254	12,303	9,602	8,354	8,156	7,891
Maximum resolution Å	2.6	2.6	3.0	3.0	3.2	3.2
Mean fraction isomorphous change§			28.4%	27.8%	16.8%	17.5%
Number of heavy atom sites			2	4	6	6
$R_c $			55.8%	50.7%	61.4%	57.9%
Correlation¶			0.47	0.59	0.43	0.49
Number of centric reflections			1,611	1,455	1,415	1,323
Overall phasing power**			1.26	1.47	1.90	1.93

EMTS is ethyl-mercuri-thiosalicylate; PHMPS is *p*-hydroxymercuri-phenyl-sulphonate.

* Data set collected on image-plate area detector at the EMBL outstation, DESY, Hamburg.

† $R_{\text{sym}} = \sum_i |I_i - \bar{I}| / \sum_i I_i$, where I_i is the intensity of the reflection and $\bar{I}$ is the mean intensity of the i observations.

‡ Number of unique reflections phased: 9,898 (15 Å–2.9 Å); overall figure of merit = 0.78.

§ Versus merged data set Native(1) + Native(2).

|| R_c is Cullis R -factor for centric reflections¹⁹.

¶ Correlation between F_{obs} and F_{calc} for heavy atom structure (program REFIN in CCP4-package¹⁸).

** Phasing power is $F_H/E_{r.m.s.}$, where F_H is heavy-atom structure factor and $E_{r.m.s.}$ is residual lack of closure¹⁹.

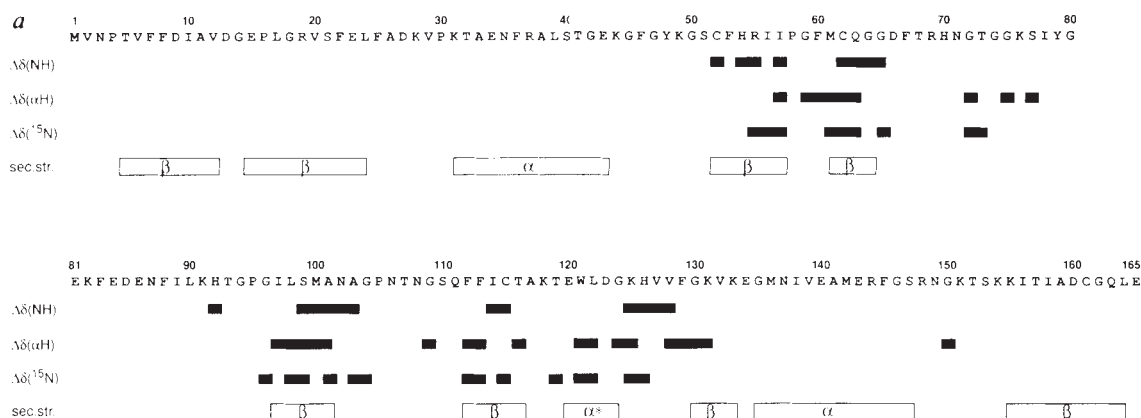
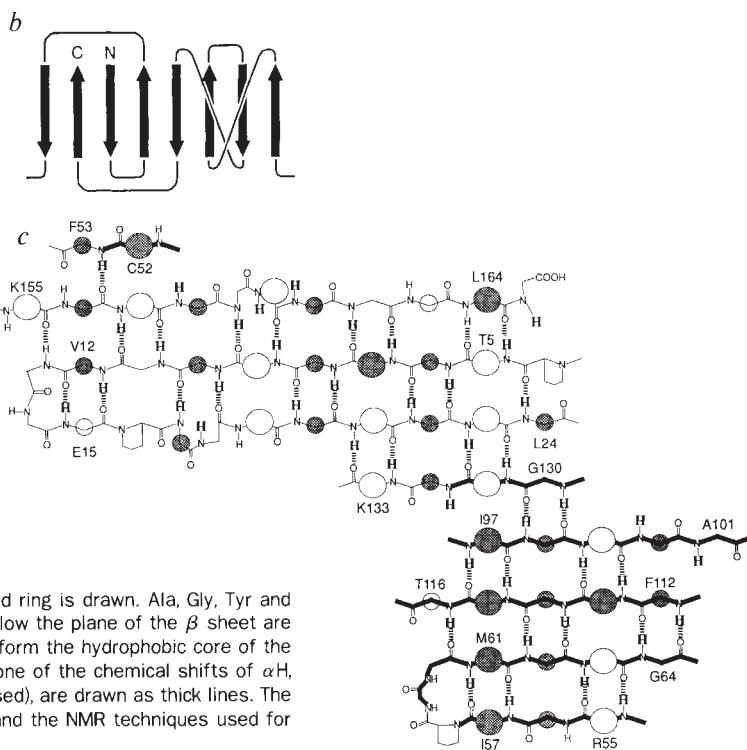


FIG. 1 Summary of the NMR results defining the CsA-binding site and the regular secondary structure and topology of cyclophilin. *a*, Amino-acid sequence of human cyclophilin, sequence locations of the regular α -helical or β -sheet secondary structures (α^* indicates that this structure could either be a short α helix or a turn-like structure) and the chemical shift data used to delineate the binding site for CsA. The residues showing significant backbone chemical shift variations on binding of cyclosporin are identified with black bars. Chemical shift differences are indicated if they exceed the following limits: $\delta(\text{NH})$, 0.10 parts per million (p.p.m.); $\delta(\alpha\text{H})$, 0.05 p.p.m.; $\delta(^{15}\text{N})$, 0.50 p.p.m. *b*, Schematic representation of the antiparallel β sheet in cyclophilin showing a global (+1, -3, -1, -2, +1, -2, -3) topology of the β sheet. *c*, Drawing showing the sequence positions of the individual β strands, the hydrogen bonds, the distribution of the hydrophobic side chains and the location of residues with large chemical shift changes on binding of cyclosporin. The two ends of each β strand are labelled with the one-letter amino-acid code and the sequence location. Slowly exchanging amide protons are printed in bold. The presence of hydrogen bonds is identified with dashed lines. The amino-acid side chains are represented with circles at the $C\alpha$ carbon positions using the following code: big circles, side chains toward the reader; small circles, side chains pointing away from the reader; shaded circles, hydrophobic residues (C, I, V, L, F, M); empty circles, polar and charged residues (S, T, N, D, Q, E, R, K). For Pro a five-membered ring is drawn. Ala, Gly, Tyr and His have no symbol at the $C\alpha$ position. Essentially all side chains below the plane of the β sheet are hydrophobic. In the X-ray structure (Figs 2 and 3) these side chains form the hydrophobic core of the β barrel. Residues that experience a significant change of at least one of the chemical shifts of αH , NH or ^{15}N after complexation with cyclosporin (see *a* for the limits used), are drawn as thick lines. The sequence-specific NMR assignments for the polypeptide backbone and the NMR techniques used for the secondary structure determination are described in ref. 7.



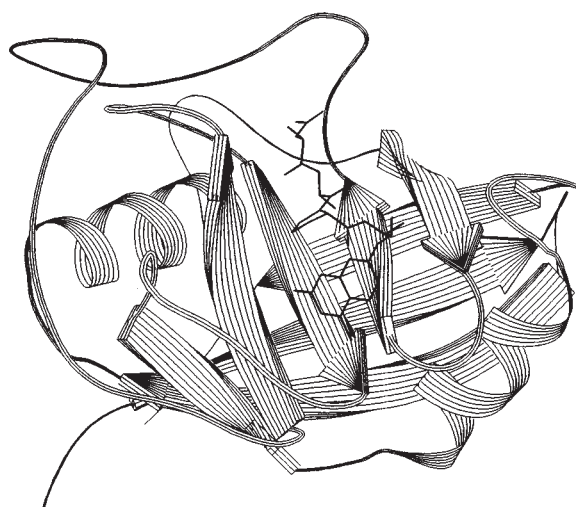


FIG. 2 A ribbon plot¹⁸ illustrating the overall fold of human cyclophilin. The model tetrapeptide prolyl isomerase substrate (*N*-acetyl-Ala-Ala-Pro-Ala-amidomethylcoumarin) is also shown. The crystal structure was solved for a complex of human recombinant cyclophilin with this linear tetrapeptide using phases determined with the multiple isomorphous replacement (MIR) method from two heavy-atom derivatives (Table 1). There are two molecules in the asymmetric unit related to each other by a noncrystallographic twofold rotation axis. The MIR phases provided an electron density map^{18,19} good enough for chain tracing²⁰ for all amino acids of both molecules. No solvent-flattening procedures were used. The secondary structure of cyclophilin as determined by NMR (ref. 7) was of considerable help in the chain tracing and model-building. The two independent molecules differ only slightly in loop regions involved in lattice contacts; the C α atoms currently have an r.m.s. fit of 0.8 Å. Noncrystallographic symmetry was not imposed during refinement using X-PLOR (ref. 21). The present *R*-factor using data between 2.5 Å and 8 Å is 25.7%. Water molecules have not yet been included and further refinement is underway.

threonine and L- α -aminobutyric acid, respectively. The three-dimensional structure of cyclophilin-bound CsA has recently been determined by nuclear magnetic resonance (NMR) (refs 5, 6). Complete sequence-specific ¹H and ¹⁵N NMR assignments for the cyclophilin backbone have also been determined and used to define the regular secondary structure⁷, which consists of two well defined α helices and an eight-stranded antiparallel β sheet (Fig. 1b).

Various crystal forms of human cyclophilin complexed with either a tetrapeptide or CsA have been grown⁸. The crystal complex of cyclophilin with CsA has six molecules in the asymmetric unit and is being studied. The legend for Fig. 2 contains details of the X-ray crystallographic structure determination of a complex of human recombinant cyclophilin with the tetrapeptide *N*-acetyl-Ala-Ala-Pro-Ala-amidomethylcoumarin. This peptide is a model substrate for the prolyl isomerase, which catalyses the *cis-trans* conversion of the alanyl-prolyl amide bond.

Cyclophilin is a roughly spherical molecule with a radius of about 17 Å (Fig. 2). The main structural feature is the eight-stranded antiparallel β barrel that has a +1, -3, -1, -2, +1, -2, -3 topology (Fig. 1b). The barrel consists of two roughly perpendicular four-stranded β sheets (Fig. 1c) connected by short junctions at residues Leu98-Gly130 and Phe53-Ile156. Inside the barrel, a tightly packed core contains most of the hydrophobic side chains (Fig. 1c). Other hydrophobic residues are located in the contact region of the two amphipathic helices with the β barrel and in the cyclosporin-binding site.

There is a structural resemblance between cyclophilin and the superfamily of proteins involved in ligand transport including retinol-binding protein (RBP), bilin-binding protein and β -lactoglobulin⁹. Most of these molecules encapsulate their ligand in the β -barrel core. By contrast, the barrel core in cyclophilin is tightly packed with hydrophobic residues (Fig. 1c), and the putative ligand-binding site is on the outside of the barrel. The topology of cyclophilin also differs from the simple (+1)_n up-and-down fold found in the RBP class of proteins, or the (-3, +1, +1) Greek key topology that is most frequently found in antiparallel β -barrel proteins. In particular the two crossover connections Gly64-Ile97 and Thr116-Gly130 represent an uncommon topological feature. As both loops lie on the outside of the barrel, the [-2, +1, -2] topology requires that the two loops cross each other. The unusual left-handed connection of the 116-130 loop may be rationalized by the length of the loop which may accommodate a variety of local conformations.

Sequence-specific NMR assignments for the ¹H and ¹⁵N spins of the polypeptide backbone of cyclophilin were also obtained for a complex formed with a water-soluble CsA derivative. The very similar nuclear Overhauser effect (NOE) patterns observed for free and CsA-bound cyclophilin indicate that there is no change in secondary structure on CsA binding. Nonetheless, comparison with free cyclophilin showed that there are residues for which cyclosporin binding caused significant chemical shift changes (Fig. 1a). These chemical shift data were used to delineate the CsA-binding site in the three-dimensional cyclophilin structure (Fig. 3).

When the cyclophilin residues thus identified as being involved in cyclosporin binding (Fig. 1a) are mapped onto the three-dimensional structure, they are found to cluster on one side of the molecule and incorporate the tetrapeptide-binding site (Figs 1c, 2 and 3). In the crystal structure, the tetrapeptide ligand binds in a long deep groove located on the protein surface between one face of the β barrel and the Thr116-Gly130 loop. The best fit in the current model shows the Ala-Pro amide bond in the *trans* conformation. Site-directed mutagenesis studies^{10,11} have been used to study enzymatic activity. Both Cys115 and Cys62 are near the peptide-binding groove, however replacing each Cys individually by Ala did not affect prolyl isomerase activity¹⁰. Site-directed mutagenesis studies have also shown that the sole tryptophan (Trp121), which sits in the middle of the binding loop, is implicated in CsA binding, but has little effect on prolyl isomerase activity¹¹. In the crystal structure, this

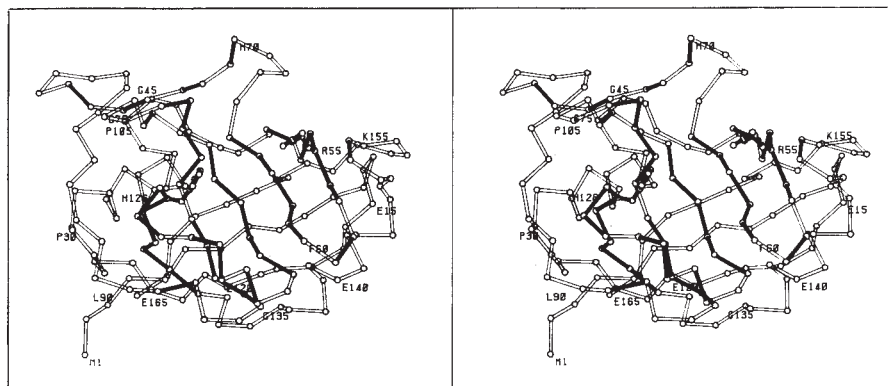


FIG. 3 Stereo PLUTO plot¹⁸ showing the α skeleton of cyclophilin. Selected C α atoms have been labelled with the sequence number and the one-letter amino-acid code. Residues which are implicated in CsA binding based on the chemical shift data in Fig. 1a, are shown in this diagram with filled C α -C α bonds. The side chains of Arg55 and His126, which may be involved in the prolyl isomerase mechanism, are also shown.

tryptophan is close to the coumarin ring of the tetrapeptide, and intermolecular NOEs observed in the cyclophilin-CsA complex in solution showed that it is in close contact with residues 9 and 11 of CsA (ref. 5). His126 sits on the same binding loop as Trp121, with N^δ about 6 Å from the carbonyl carbon of the prolyl amide. A possible mechanism for prolyl isomerase activity could involve His126 acting as proton acceptor for a water molecule involved in nucleophilic attack on the carbonyl carbon of the Ala-Pro amide bond. Further activation of the amide carbonyl group could be provided by the hydrogen bond between the carbonyl oxygen atom and the guanidinium group of Arg55.

The macrolide FK506 is chemically unrelated to CsA but is also a potent immunosuppressant with a very similar biological profile¹². FK506 binds to the specific cytosolic immunophilin protein receptor FK-binding protein (FKBP) (ref. 12). Human recombinant FKBP has a chain length of 107 amino acids (M_r , 11.8K) and has no significant sequence homology with cyclophilin. The recently determined X-ray and NMR structures of human recombinant FKBP (refs 13–15) show that it exists in the crystal and in solution as a five-stranded antiparallel β -sheet which wraps around a short helix. There is thus no readily apparent similarity with the three-dimensional cyclophilin structure. The FK506 class of drugs do not bind to cyclophilin and CsA does not bind to FKBP. Nonetheless the mode of action of CsA and FK506 in the cell seem to be very similar^{12,16}. FKBP also shows a *cis-trans* prolyl isomerase activity. But for both FK506 and CsA, suppression of prolyl isomerase activity alone seems insufficient to explain the biological activity, as drug concentrations necessary to inhibit T-cell activation would not saturate the abundant prolyl isomerases^{12,17}.

Our results are in line with the earlier NMR reports on CsA bound to cyclophilin⁷: intermolecular NOEs determined between cyclosporin and CsA implicate residues along one face of cyclosporin (residues 9, 10, 11, 1 and 2) as being important for binding. These complementary structural studies using both NMR spectroscopy in solution and X-ray crystallography thus provide new insights into the binding of CsA and, more generally, the molecular basis of immunosuppressive activity.

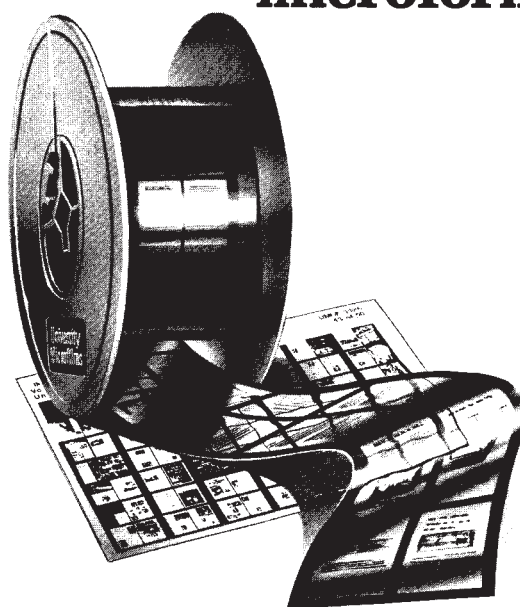
Note added in proof: The current *R*-factor is 21.8% for all data from 8 to 2.3 Å. An independent X-ray structure of unliganded cyclophilin has been determined²². Despite different unit cells, the molecular architecture seems similar. □

Received 5 July; accepted 1 August 1991.

1. Handschumacher, R. E., Harding, M. W., Rice, J., Drugge, R. J. & Speicher, D. W. *Science* **226**, 544–547 (1984).
2. Borel, J. F. *Pharmac. Rev.* **41**, 259–371 (1989).
3. Takahashi, N., Hayano, T. & Suzuki, M. *Nature* **337**, 473–475 (1989).
4. Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F. X. *Nature* **337**, 476–478 (1989).
5. Weber, C. *et al. Biochemistry* **30**, 6563–6574 (1991).
6. Fesik, S. W. *et al. Biochemistry* **30**, 6574–6583 (1991).
7. Wüthrich, K., Spitzfaden, C., Memmert, K., Widmer, H. & Wider, G. *FEBS Lett.* (in the press).
8. Zurini, M. *et al. FEBS Lett.* **276**, 63–66 (1990).
9. Cowan, S. W., Newcomer, M. E. & Jones, T. A. *Proteins* **8**, 44–61 (1990).
10. Liu, J., Albers, M. W., Chen, C., Schreiber, S. L. & Walsh, C. T. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2304–2308 (1990).
11. Liu, J., Chen, C. & Walsh, C. T. *Biochemistry* **30**, 2306–2310 (1991).
12. Schreiber, S. L. *Science* **251**, 283–287 (1991).
13. Michnick, S. W., Rosen, M. K., Wandless, T. J., Karplus, M. & Schreiber, S. L. *Science* **252**, 836–839 (1991).
14. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. *Science* **252**, 839–842 (1991).
15. Moore, J. M., Peattie, D. A., Fitzgibbon, M. J. & Thomson, J. A. *Nature* **351**, 248–250 (1991).
16. Tropschug, M., Barthelmess, I. B. & Neupert, W. *Nature* **342**, 953–955 (1989).
17. Blierer, B. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 9231–9235 (1990).
18. CCP4; A Suite of Programs for Protein Crystallography (SERC Daresbury Laboratory, Warrington, UK, 1986).
19. Blundell, T. L. & Johnson, L. N. *Protein Crystallography* 357 (Academic, Oxford, 1976).
20. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
21. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
22. Ke, H. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).

ACKNOWLEDGEMENTS. We thank R. Wenger for supplying the water-soluble cyclosporin and K. Wilson and C. Betzel of the EMBL outstation, Hamburg, for their help with data collection using synchrotron radiation. This work was supported by the Schweizerischer Nationalfonds (to K.W.).

nature is available in microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

University
Microfilms
International

One-stop shopping

US patents documents on CD-ROM, a personal densitometer, a beta counter that eliminates the need for liquid radioactive waste disposal, and a blot visualization spray — what's new and what's hot in the scientific product line.

FOR up-to-date access of US patent documents, MicroPatent offers Patent-Images, a fully searchable **data base of complete facsimile US patents on CD-ROM**, which is supplied with software specially developed for patent information (*Reader Service No. 101*). Delivered



US patent information data base.

weekly, each PatentImages disk contains complete information on up to 1,000 patents only three weeks after their publication by the US Patent Office. Patent-Images provides users with full electronic access to documents that previously were available only in printed form or on microfilm. PatentImages may be of particular interest to people who require access to patent diagrams, drawings or structures. Using Patsoft software, which was developed specifically for patent information by the European Patent Office, users can access any patent via 12 search fields; this is designed to allow fast retrieval of the latest patents. Having located a patent, users can rotate and zoom in on the image, or look at double-page spreads on screen. Each disk contains an index of the year's patents for the purposes of cross-referencing. Free evaluation disks are available.

The PPM Star Catalogue, which can be obtained through Spektrum Akademischer Verlag, is an **astrometric reference star catalogue** that uses the new IAU coordinate system of FK5 (J2000) (*Reader Service No. 102*). Designed to replace the SAO catalogue and AGK3 as astronomic references, the PPM Star Catalogue provides more than 180,000 reference stars on the northern celestial hemisphere and has a much higher accuracy than the two older catalogues, the company says. Both the mean errors of the proper motions and of the present epoch positions provide an improvement

of more than a factor of 3 as compared to the SAO catalogue, says Spektrum. The star density is increased by 35 per cent. In addition to the astrometric data (positions, proper motions and their mean errors), the following data are given for each star: photographic magnitude, spectral type, HD number, SAO number, AGK3 number, Durchmusterung number and some remarks. The physical units used for the data and the arrangements of the stars are the same as those used in the SAO catalogue. The final version of the southern hemisphere complement to PPM will be available by year end.

The **VIDEOPLAN image analysis system** from Kontron Elektronik provides a variety of interactive measuring methods for complex images (*Reader Service No. 103*). Users can derive a complex range of morphological and stereological parameters quickly and accurately from any image that is available as either a video signal (VCR, black and white or



Kontron's image analysis system.

colour camera) or hard copy, the company says. The measurement sequence and structure can be customized using set-up files, parameter definition and arithmetics (online during the measurement or off-line once all the features have been measured) to meet user-specific requirements. In addition, the high-precision digitiser tablet ensures the accuracy of measurement results. A comprehensive statistics package provides a host of options for data evaluation, including histograms, scattergrams, Kolmogoroff-Smirnoff Test, χ^2 -Test, T-Test, G-Test and H-Test. An optional three-dimensional reconstruction package is also available.

Video view

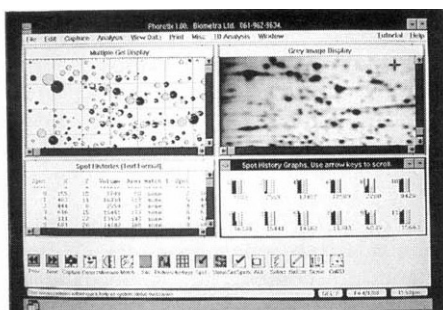
Occupying only 48 cm of bench space, the new Personal Densitometer from Molecular Dynamics is designed to provide a



Laser-based densitometer for films, gels, blots and TLC plates.

complete, fast-scanning, **laser-based densitometer** (*Reader Service No. 104*). The optical system, which uses a pentaprism to scan rapidly the laser beam across the sample, allows the Personal Densitometer to scan films, wet and dry gels, membrane blots and thin-layer chromatography plates by transmission densitometry. Background variations are eliminated using a beam splitter in the optical path and control circuitry that maintains constant energy output from the laser. The combination of the light integrating cylinder and transmission densitometry provides linear performance from 0 to 4 optical density, Molecular Dynamics says. In addition, high-resolution scans at 50- μ m resolution provide images that are suitable for even the most demanding applications such as two-dimensional protein gels, the company says. The 4,096 levels of optical density resolution ensure that even the faintest bands can be seen and quantified. The Personal Densitometer also provides an 8-bit image format, which is compatible with most Apple Macintosh image processing software. It also features the company's Image Quant image analysis and quantification software, which runs on the built-in AT 386SX computer. Standard with each Personal Densitometer is the advanced VGA display that is designed to provide sharp, clear images in 256 colours and 64 gray levels.

Phoretix-1, a system for the semi-automatic **quantitation and analysis of**



Phoretix-1 for 2-D gel analysis.

two-dimensional gels, is now available from Biometra (*Reader Service No. 105*). The system, which runs under Microsoft Windows Version 3, allows the user to measure and trace accurately the behaviour of an unlimited number of proteins through up to 36 gels per experiment, Biometra says. The system inputs images from a video camera but an interface to scanners, densitometers or higher-resolution cameras can be provided, as required. All spots can be individually tagged to facilitate comparisons between experiments. One of the main benefits of the system is the ability to select proteins of interest for analysis, either manually, or by the use of parameters. This could be useful in experiments that generate large numbers of proteins. All results can be viewed or printed in graphic or text format, or can be output to spreadsheets, statistics packages or desktop publishing packages for report generation. Other benefits include non-interlaced video output to eliminate eye-strain, extensive editing facilities and the ability to run the system either interactively or in batch mode. The software will run on any '386 or '486 PC or compatible computer.

Counting devices

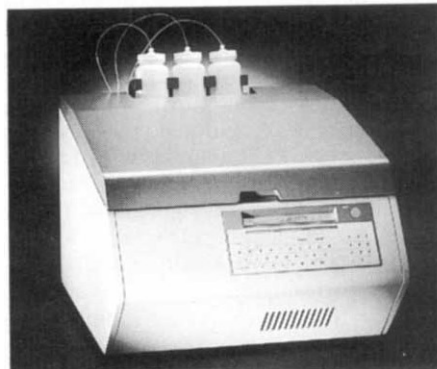
Packard's Matrix 9600 Direct Beta Counter automatically counts radio-labelled solid supports in the 8×12 microplate format; this dispenses with the need for vials and liquid scintillation cocktail and, hence, eliminates the need to dispose of liquid radioactive waste (*Reader Service No. 106*). The **beta counter** accepts up to 25 filter plates at a time, which provides a throughput of up to 2,400 samples an hour (based on a counting time of two minutes). The company claims that for cell proliferation and receptor binding assays, and dot blot



Cocktail-free beta counter.

analyses, the Matrix 9,600 provides results two to forty times faster than single- and multi-detector liquid scintillation and gamma counters.

Temperature control, three injectors, a 150-sample capacity and high sensitivity are features of the new LB953 Autolumat **automatic luminometer** from E G & G Berthold (*Reader Service No. 107*). Users can set the temperature of the sample chamber from 5 °C above ambient to 45 °C. To ensure low background (less than 100 c.p.s.), the broad-spectrum photomultiplier tube, which has a range of 370–620 nm, is cooled to 6 °C, the company says. The incubated sample chamber can hold up to 150 samples. The instrument can store up to 60 user-defined protocols and can be controlled through the front-panel display or via a computer. For kinetic measurements, the



E G & G's automatic luminometer.

instrument can be set to measure repeatedly each sample over time. Between measurements, the samples are kept in the temperature-controlled incubation area. Optional software for use with the LB953 provide a range of data evaluation routines.

Link up

Three new sizes of **hybridization cassette** for northern and Southern blotting have been added to the TIMEFRAME equipment line from Bios (*Reader Service No. 108*). In addition to the standard (9.3×22 cm) and large (22×22 cm) cassettes, Bios offers 11×16 cm, 12×25 cm and 22×22 cm cassettes. All five can be used for membrane hybridizations in either an oven or water bath and require only 20–25 per cent of the amount of probe required in standard Southern and northern procedures. In addition, the reusable cassettes eliminate the need for heat-sealed bags and minimize the risk of coming into contact with radioactive material, Bios says. The membrane is placed in the cassette and prehybridized. After prehybridization, the prehybridization buffer is shaken off the membrane and the hybridization solution layered onto the membrane. A plastic cover keeps the solution in contact with the membrane.

The cassette is then closed and incubated at the appropriate temperature. Hybridization solution is removed through the stopcocks and all the washing steps are performed in the cassette itself.

Bio-Rad has introduced the GS Gene Linker **UV chamber**, which is designed to crosslink DNA or RNA to either charged or neutral nylon membranes in less than two minutes (*Reader Service No. 109*). This can, Bio-Rad says, represent considerable time savings as the conventional method of baking nucleic acids to a membrane can take up to an hour. Crosslinking with the GS Gene Linker is designed to enhance multiple probings of Southern, northern, slot and dot blots, and colony or plaque screening. Other uses for the instrument include nicking DNA in agarose gels before transfer, UV sensitivity testing of host strains, UV sterilization and UV crosslinking of proteins to nucleic acids. To facilitate use of the GS Gene Linker, the instrument has been preprogrammed with the optional energy settings for most of the commonly used applications. As the photodetector system measures only the usable energy emitted from the lamps, turning them off after the desired energy has been delivered, reproducible results are ensured, Bio-Rad says. For users requiring optimal signal detection, Bio-Rad supplies an optional 312-nm conversion kit. In addition, the photodetector can quickly be recalibrated by the user for use with 312-nm bulbs.

Detection systems

Visiblot AP is the new ready-to-use alkaline phosphatase **blot visualization spray** from United States Biochemical, which produces a semi-permanent dark-purple spot or band on the membrane or solid support at the alkaline phosphatase binding site (*Reader Service No. 110*). It can be used in a variety of assays for either quantitative or qualitative analysis, including protein, DNA or RNA slot or dot blots, western, Southern or northern blots, coated beads or dipsticks, flow-through sample assay chambers, flow-sorted cells or chromosome spot blots. USB claims that Visiblot AP is as sensitive as BCIP/NBT and produces reduced background staining. The substrate is supplied as a non-aerosol spray formulation, which is designed to eliminate waste and the potential for pipette contamination. It is compatible with the company's Protein Images and Gene Images non-isotopic detection kits. Blots first may be sprayed with the Visiblot AP reagent and then subsequently detected with Lumi-Phos chemiluminescence reagent and X-ray film without the need for intermediate treatment. Visiblot AP is sold in 125-ml quantities for \$85 (US) and is stable for over a year at 45 °C when kept

protected from light.

AmpliProbe is the new **nonradioactive DNA/RNA probe system** from Imclone Systems, which takes the user from hybridization to results in less than seven hours (*Reader Service No. 111*). Results, which are recorded on standard X-ray film, can be obtained from northern, Southern and



AmpliProbe's nonradioactive DNA/RNA probe system.

slot blots, and colony lifts, says Imclone. AmpliProbe is available in target-specific kits, as well as a universal detection kit containing DNA that can be modified by the user so as to act as a probe for a target of choice. The initial line of target-specific kits include probes for five viruses (cytomegalovirus, Epstein Barr virus, hepatitis B virus, human immunodeficiency virus and herpes simplex virus 1 and 2), three oncogenes (*c-myc*, *n-myc*, *erb B-2*), and a human beta-actin. The patented signal amplification technology used in the AmpliProbe kits is designed to concentrate as much signal as possible on the target sequence. Imclone says, this is achieved with a series of synthetic probes — the AmpliProbe secondary probes — which have been designed specifically to be complementary to the backbone of M13, a widely used vector into which the target sequences are cloned. Each secondary probe is conjugated to alkaline phosphatase, the readout moiety. Multiple secondary probes hybridize to the M13 backbone, resulting in hybridization of multiple readout moieties to the target, Imclone says. Sensitivity levels comparable to the use of radioactive isotopes can be achieved with the kits.

Molecular miscellany

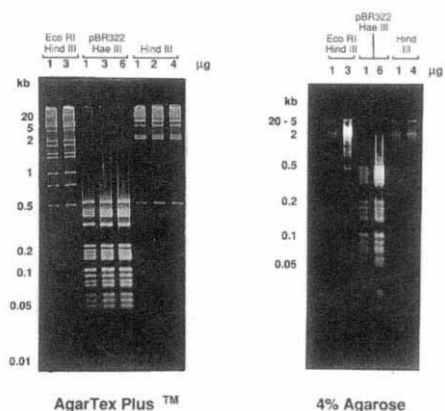
For the **one-step isolation of double-stranded nucleic acids** from mixtures containing buffers, salts, nucleotides and primers, Advanced Genetic Technologies recommends its DS SELECT kit (*Reader Service No. 112*). Using DS SELECT, purified double-stranded product can be recovered from aqueous solution in less than three minutes, the company says. The kit, which is suited to the rapid processing of multiple samples, is designed to recover greater than 95 per cent of double-stranded DNA and remove greater than 99 per cent of primers. Researchers may find DS SELECT use-

ful for applications such as the polymerase chain reaction, sequencing and solution hybridization. DNA purified using DS SELECT is suitable for electrophoretic analysis, restriction mapping and cloning purposes.

Clontech introduces QUICK-Clone cDNA, **double-stranded, high-purity cDNA** that is ready for PCR amplification (*Reader Service No. 113*). Use of QUICK-Clone cDNA offers researchers added convenience in isolating and cloning genes as the need to purify RNA or make cDNA libraries is eliminated. Researchers use their own primers with QUICK-Clone cDNA to amplify and clone genes of interest, rapidly acquire cDNA hybridization probes, or to determine expressed gene regions. Over 25 QUICK-Clone cDNAs are available for several human, mouse and rat tissues, including brain, liver, lung, heart, spleen and placenta, as well as for CHO cells. QUICK-Clone cDNA is synthesized from Poly A⁺ RNA using an oligo(dT) primer. It is then purified to remove interfering RNA and genomic DNA for more specific PCR amplification. In addition, QUICK-Clone cDNA is size-selected to ensure a high probability of isolating full-length cDNAs, Clontech says. Each QUICK-Clone cDNA includes two aliquots of cDNA (approximately 10 ng each) and an applications guide.

Gel assortment

AgarTex Plus is the new **synthetic gel matrix** from National Diagnostics that is designed for the separation of PCR fragments and the recovery of samples that can be used immediately without the need for further purification procedures (*Reader Service No. 114*). The gel, which is intended as an agarose replacement for electrophoretic separations of double-stranded DNA fragments from 8 to 20,000 base pairs, is free of the contaminants found in agarose that inhibit the activity of certain enzymes. In addition,



AgarTex Plus — National Diagnostics' synthetic gel matrix and agarose substitute.

AgarTex Plus gel can accommodate ten times more sample than agarose gels, the company says. The composition of AgarTex Plus confers several unique properties: it melts at low temperatures under mild conditions, gels only once taking about an hour and, upon heating, will melt and remain liquified. In addition, the AgarTex Plus matrix will not interfere with subsequent reactions or enzymatic procedures, the company says. Although AgarTex Plus is optimized as a replacement for four per cent agarose in the separation of fragments generated from PCR procedures, it also can replace one per cent agarose in the separation of larger DNA fragments. AgarTex Plus is supplied in 500-ml and 1-litre amounts and can be disposed of as non-toxic, non-hazardous waste.

A new inexpensive alternative to CNBr (cyanogen bromide) agarose called Triazine-Activated 4 × L has been developed by Affinity Chromatography (*Reader Service No. 115*). The **agarose** utilises a novel triazine bonding chemistry which is highly reactive with proteins, ACL says. Antibodies or enzymes can be bonded to the crosslinked agarose sup-



Triazine-activated agarose.

port under relatively mild conditions. High coupling efficiencies minimize loss of activity and wastage of valuable protein, ACL says. Triazine chemistry has several advantages over cyanogen bromide coupling, including high reactivity at near neutral pH and the generation of a stable neutral bond. Triazine-activated agarose is supplied as a freeze-dried powder in a variety of pack sizes.

Reagent round-up

Through British Bio-technology, researchers now have access to a selection of **tools for investigating the role of interleukin 8 (IL-8)**, an 8-kD protein that demonstrates a range of activities, including the stimulation of chemotaxis in neutrophils, basophils and T lymphocytes (*Reader Service No. 116*). The new product range includes an ELISA system,



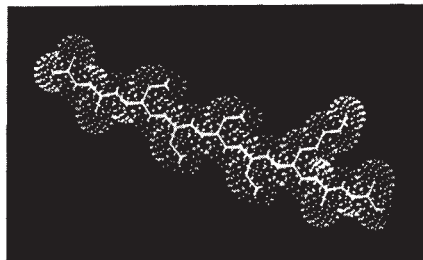
British Bio-technology's IL-8 product line.

the recombinant protein and a polyclonal antibody. The ELISA has a maximum sensitivity of 4.7 pg ml^{-1} and can be used to measure human IL-8 in tissue culture supernatants, sera, plasma and other biological fluids. False-negative results due to complement activation and false-positive results caused by the presence of heterophilic antibodies are prevented by reagents included in the kit, the company says. The human protein is produced in *Escherichia coli* from the company's Designer Gene for IL-8, and can be supplied with or without carrier protein in 10- and 50- μg quantities. Completing the line-up, the polyclonal antibody is raised in goats and can be used in applications that include neutralising bio-assays, western blots and ELISAs.

Laccase, an **enzyme** produced from the growth media of *Coriolus hirsutus*, a wood-attacking fungus, is now available from C-C Biotech (Reader Service No. 117). The company claims that laccase provides an attractive alternative to horseradish peroxidase: it has no endogenous activity and is not inhibited by substrates. Laccase catalyses the reaction of hydroxyquinone to *o*-benzoquinone and pyrocatechol to *p*-benzoquinone. It will also oxidize ascorbic acid, *p*-cresol, *o*-phenylenediamine, ABTS and other substrates. The activity of laccase can be determined potentiometrically, electrochemically and, most commonly, spectrophotometrically. Laccase produces a colour reaction with most peroxidase substrates and uses oxygen instead of hydrogen peroxide as the second substrate. The enzyme can, for example, be used in all types of immunoassays and in peptide synthesis for the quick removal of phenyl hydrazide protection groups under mild conditions. The enzyme is supplied as a freeze-dried preparation from 0.01 M sodium phosphate buffer. In this form, C-C Biotech says it is stable for at least two years without loss of activity.

Ultrapure **enterokinase** is now available from Biozyme (Reader Service No. 118). This specific protease has been highly purified for the cleavage of fusion

proteins containing the enterokinase recognition sequence. As Biozyme's ultrapure enterokinase offers high specific activity (100,000 units per milligram of protein, trypsinogen as substrate, 25 °C, pH 5.6), it can be used in minimal

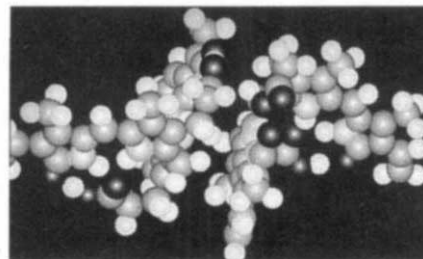


Enterokinase recognition sequence.

quantities; this reduces the inherent risk of product degradation, the company says. Biozyme can supply the enzyme in quantities to suit research or production requirements.

Applications software

Version 3.0 of Ball and Stick, the **three-dimensional molecular graphics and presentation package** from Cherwell Scientific, is now available for Macintosh computers (Reader Service No. 119). This research and teaching tool includes a host of features for importing, viewing, editing and outputting publication-quality molecular models. Working within the familiar Macintosh menu-driven environment, Version 3.0 of Ball and Stick features: one-step commands for the importation of coordinates from a wide variety of structural modelling packages; a vast array of display options to customize virtually every aspect of the molecule — five different display styles, stereo

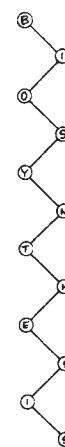


Ball and Stick Version 3.0 for molecular modelling on Macs.

images, atom labels that can be customized, tonal shading, zooming and clipping; animations of rotations; two concurrent display windows; and the facility to export to other word processing and drawing packages. This latest version of Ball and Stick now provides real-time interactive viewing. Using the mouse, molecules can be placed anywhere in three-dimensional space and viewed from any direction. Real-time operations include rotation, XY translation, XZ translation, distance change, angle change and torsion change. Version 3.0 sells for £249 (UK) (to commercial users) and £149 (UK) (to educational users).

ADVERTISEMENTS

PEPTIDES Around the World



- Two weeks turnaround
- \$65.00/residue, 200 μmole scale
- Amino acid analysis included
- Experience and reliability you can count on

Other Services:
Carrier Conjugation
Antiserum Production
Protein-Peptide Sequencing

1-800-227-0627

Your Source of Peptides

BIO-SYNTHESIS

Fax 1-214-420-0442

Reader Service No.18

Custom Peptides

\$30 a residue, 10 μmole scale
\$80 a residue, 100 μmole scale

Research Genetics

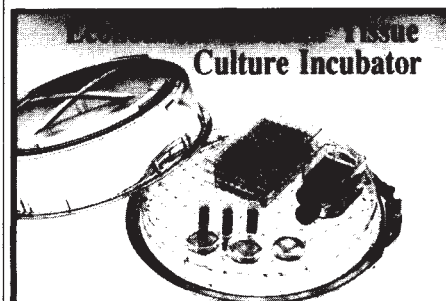
1-800-533-4363

In the United Kingdom 0 800 89-1393

In Canada 1-800-533-4363

FAX 205-536-9016

Reader Service No.32



- Safe for AIDS research • Reliable for In-Vitro fertilization
- Versatile enough for thousands of lab uses

Easy to use — Just flush with gas mixture, seal and place at approp. temp. Perform multiple experiments simultaneously w/out cross contamination. Thousands in use worldwide.

For information contact: **Flow Laboratories, Ltd.** or, **Billups-Rothenberg, Inc.** • P.O. Box 977
Del Mar, CA 92014-0977 USA • (619) 755-3309

Reader Service No.37

Jandel Scientific announces the release of SigmaPlot 4.1, a major upgrade of the **scientific graphics software** for IBM PCs and compatibles that can be used to create publication-quality charts and graphs (*Reader Service No. 120*). The new version includes changes to facilitate use, reduction in required memory size, utilization of expanded and extended memory capabilities, and increased font and output support. Easy-to-use 'drop-down' menus are a new feature of version 4.1, Jandel says. An 'information bar' has been added at the bottom of the screen, which offers the user feedback information on selected options. In addition, SigmaPlot 4.1 features a new 'dynamic memory' system that uses smaller memory overlays and brings into memory only what is needed. As a result of this, version 4.1 uses over 40 K less memory than the earlier version and runs concurrently with memory-resident programs. Another useful feature is SigmaPlot's user-defined system defaults. This lets the user define graph attributes (type of plot, line thickness, colours, font choice), which can be saved and reused in future graphs. The user chooses 'Set Defaults' to save the current graph's settings as new defaults for future graphs. SigmaPlot 4.1 costs DM1,150 (Germany).

ADVERTISEMENT

Synthetic Oligonucleotides

Custom made probes and primers,
synthesised, purified and shipped in 3 days

£3.50 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

Modified bases, biotin or digoxigenin labelling,
data base analysis and full design facilities

Molecular Medicine Unit,

King's College School of Medicine & Dentistry,
Denmark Hill, London SE5 8RX

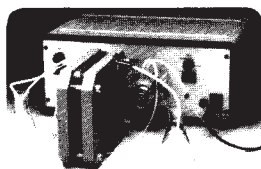
Tel. XX-44-(0)71 - 326 3126

FAX XX-44-(0)71 - 733 3877

Reader Service No.56

LIPOSOMES

REPRODUCIBLE - HOMOGENEOUS -
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

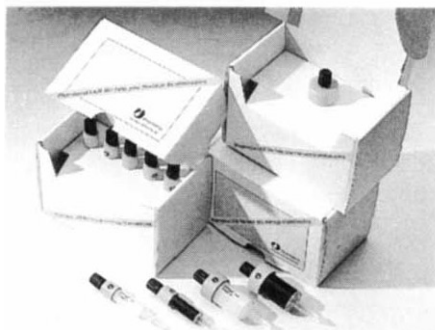
DIABORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.17

Odds and ends

Pharmacia has introduced a new range of Hi Trap pre-packed, multiple-use **affinity chromatography columns** (*Reader Service No. 121*). Protein A and Protein G media are included in the HiTrap range of six affinity gels. All are based on Sepharose High Performance, a stable base matrix for fast and convenient separations, Pharmacia says. The full range consists of five



HiTrap affinity columns.

ready-coupled affinity gels (HiTrap Protein A, HiTrap Protein G, HiTrap Heparin, HiTrap Blue and HiTrap Chelating) and one pre-activated gel (HiTrap NHS-activated) onto which you can couple your own ligand. Each HiTrap gel is pre-packed and ready-to-use in tough, transparent 1-ml and 5-ml plastic columns with syringe and pump connections. Detailed instructions are also included.

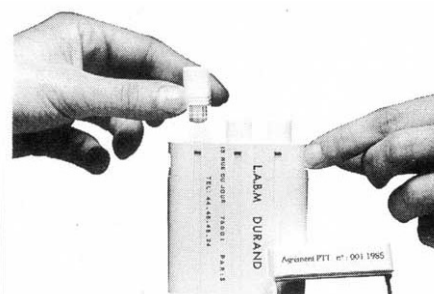
The PCR/Sequencing Kit from Applied Biosystems can be used for the **direct amplification and automated sequencing of target DNA**, ranging from



genomic to plasmid DNA (*Reader Service No. 122*). The chemistry and protocols are based on symmetric GeneAmp PCR and Taq cycle sequencing chemistry. Researchers who use the kit no longer have to clone before selecting and amplifying enough target DNA for automated base calling on the 373A DNA Sequencer. PCR amplification replaces cloning completely for many applications, including the sequencing of inaccessible samples such as human immunodeficiency virus, the company says. Double-stranded sequencing from either strand can be performed directly from the symmetric PCR fragment. Symmetric PCR requires no titration of primer concentra-

tions. Because the procedure is robust, there is no need for template-to-template optimization.

The French company Samedex has developed a new **container for the safe transportation of biological samples** such as blood (*Reader Service No. 123*). The use of shock-resistant materials and a system for capping and securing the tubes



Shock-resistant container for the safe transportation of biological samples.

in the container to provide a watertight seal are design features of the container. Made of fire-resistant thermoplastic, the container has three cylindrical compartments that hold the sample tubes. The tubes are made of crystal polycarbonate, a material that combines the advantages of metal, glass and thermoplastics, Samedex says. It retains its high shock resistance over a temperature range from -30 °C to 135 °C and keeps its shape at temperatures as high as 146 °C. The tube caps are made from high-density, high-pressure polyethylene. When clipped on to the container, the container's cover holds sample tubes firmly in position. A foam block helps to absorb any shocks. The system was designed to ensure complete safety. In the rare event of a tube breaking, its contents will be confined to the tube's individual compartment without the risk of outside contamination, the company says. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

CORRECTION

With reference to the product review (*Nature* 351, 676, 1991) on a yeast artificial chromosome (YAC) library constructed at ICI Pharmaceuticals, *Nature* wishes to apologise for the implication that ICI Pharmaceuticals has set up a commercial screening service for their library. In fact, ICI management, after discussions with the Medical Research Council (MRC), donated a copy of their library to the MRC Human Genome Mapping Programme (HGMP) Resource Centre for the benefit of the research community. For further information about this library please contact MRC HGMP Resource Centre, Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, United Kingdom.